STATISTICS IN PSYCHOLOGY AND EDUCATION

HENRY E. GARRETT


EX-
LIBRIS
AMERICAN FOUNDATION
FOR THE BLIND INC.

Digitized by the Internet Archive
in 2012 with funding from
Lyrasis Members and Sloan Foundation

http://www.archive.org/details/statisticsinpsyc00henr

STATISTICS IN PSYCHOLOGY AND EDUCATION

STATISTICS IN PSYCHOLOGY AND EDUCATION

BY

HENRY E. GARRETT

ASSISTANT PROFESSOR OF PSYCHOLOGY, COLUMBIA UNIVERSITY

WITH AN INTRODUCTION BY

R. S. WOODWORTH

PROFESSOR OF PSYCHOLOGY, COLUMBIA UNIVERSITY

LONGMANS, GREEN AND CO.

55 FIFTH AVENUE, NEW YORK

CHICAGO, TORONTO, LONDON

1926

Copyright, 1926, by
LONGMANS, GREEN AND CO.

First Edition, January, 1926
Reprinted, November, 1926

HA33
G192
Cop.1

MADE IN THE UNITED STATES

INTRODUCTION

Modern problems and needs are forcing statistical methods and statistical ideas more and more to the fore. There are so many things we wish to know which cannot be discovered by a single observation, or by a single measurement. We wish to envisage the behavior of a man who, like all men, is rather a variable quantity, and must be observed repeatedly and not once for all. We wish to study the social group, composed of individuals differing one from another. We should like to be able to compare one group with another, one race with another, as well as one individual with another individual, or the individual with the norm for his age, race or class. We wish to trace the curve which pictures the growth of a child, or of a population. We wish to disentangle the interwoven factors of heredity and environment which influence the development of the individual, and to measure the similarly interwoven effects of laws, social customs and economic conditions upon public health, safety and welfare generally. Even if our statistical appetite is far from keen, we all of us should like to know enough to understand, or to withstand, the statistics that are constantly being thrown at us in print or conversation—much of it pretty bad statistics. The only cure for bad statistics is apparently more and better statistics. All in all, it certainly appears that the rudiments of sound statistical sense are coming to be an essential of a liberal education.

Now there are different orders of statisticians. There is, first in order, the mathematician who invents the method for performing a certain type of statistical job. His interest, as a mathematician, is not in the educational, social or psychological problems just alluded to, but in the problem of devising instru-

ments for handling such matters. He is the tool-maker of the statistical industry, and one good tool-maker can supply many skilled workers. The latter are quite another order of statisticians. Supply them with the mathematician's formulas, map out the procedure for them to follow, provide working charts, tables and calculating machines, and they will compute from your data the necessary averages, probable errors and correlation coefficients. Their interest, as computers, lies in the quick and accurate handling of the tools of the trade. But there is a statistician of yet another order, in between the other two. His primary interest is psychological, perhaps, or it may be educational. It is he who has selected the scientific or practical problem, who has organized his attack upon the problem in such fashion that the data obtained can be handled in some sound statistical way. He selects the statistical tools to be employed, and, when the computers have done their work, he scrutinizes the results for their bearing upon the scientific or practical problem with which he started. Such an one, in short, must have a discriminating knowledge of the kit of tools which the mathematician has handed him, as well as some skill in their actual use.

The reader of the present book will quickly discern that it is intended primarily for statisticians of the last-mentioned type. It lays out before him the tools of the trade; it explains very fully and carefully the manner of handling each tool; it affords practice in the use of each. While it has little to say of the tool-maker's art, it takes great pains to make clear the use and limitations of each tool. As any one can readily see who has tried to teach statistics to the class of students who most need to know the subject, this book is the product of a genuine teacher's experience, and is exceptionally well adapted to the student's use. To an unusual degree, it succeeds in meeting the student upon his own ground.

R. S. Woodworth

Columbia University

PREFACE

The present day emphasis on measurement and the quantitative treatment of results has made a knowledge of statistical method not only extremely useful but almost necessary to the student of psychology, education, and the social sciences. To those who have been well trained in mathematics, the acquisition of statistical technique offers no particular difficulty. To many otherwise capable students, however, either because of inadequate preparation in mathematics, or because their preparation is not very recent, the application of statistical method to data obtained from test and experiment is more than ordinarily difficult.

It is for this last group of students, especially, that this book has been written. Its primary purpose is to present the subject in a simple and concise form understandable to those who have no previous knowledge of statistical method. With this end in view, theory has everywhere been subordinated to practical application, and numerous illustrations of the various statistical devices have been provided. References have been given, however, for the benefit of those interested in the mathematical theory underlying the methods introduced.

The reader will note that in nearly all cases formulas have simply been stated without proof. This has been done, because the writer believes that most students of mental and social measurement are—and probably should be—more concerned with what a formula means and does than in how it is derived. There is considerable justification for such an attitude. In every science certain facts obtained from other fields must be taken on faith. We do not, to take a simple example, restrict the use of the radio or the microscope to those who understand the physical principles involved, and there seems to be no real

reason why a student of psychology should not make good use of a correlation formula when he cannot derive it mathematically.

A chapter has been given to the subject of reliability—a topic too often passed over lightly—and considerable space has been devoted to correlation. An entire chapter, also, has been given to partial and multiple correlation. This method, while comparatively recent, is being widely used in educational research, and is probably destined in the near future to be more often used in the psychological laboratory. In the last chapter, the application of correlation and other statistical methods is shown to tests and testing.

Many have contributed to the making of this book of whom only a few can be mentioned. To Professors R. S. Woodworth and Mark A. May who read the manuscript, the writer is indebted for many useful and constructive criticisms. He is also grateful to Dr. M. R. Neifeld, to Mr. V. W. Lemmon, and to Miss Elizabeth Farber for computations and helpful suggestions.

Henry E. Garrett

Columbia University

CONTENTS

CHAPTER I

THE FREQUENCY DISTRIBUTION

CHAPTER II

GRAPHIC METHODS AND THE NORMAL CURVE

CHAPTER III

THE RELIABILITY OF MEASURES

CHAPTER IV

CORRELATION

CHAPTER V

PARTIAL AND MULTIPLE CORRELATION

CHAPTER VI

SOME APPLICATIONS OF STATISTICAL METHOD AND TECHNIQUE TO TESTS AND TEST RESULTS

STATISTICS IN PSYCHOLOGY AND EDUCATION

CHAPTER I

THE FREQUENCY DISTRIBUTION

I. The Tabulation of Measures into a Frequency Distribution

1. Measures in General: Continuous and Discrete Series

In the measurement of mental and social traits or capacities most of the facts with which we deal fall into what are known as continuous series. A continuous series may be defined simply as a series which is theoretically capable of any degree of subdivision. *IQ*'s, for example, are generally thought of as increasing by increments of 1 on a scale which extends from the idiot to the genius; however, there is actually no real reason—at least theoretically—why with more refined methods of measurement we should not be able to get *IQ*'s of 100.8 or even 100.83. Nearly all capacities measured by mental and educational tests and scales, as well as such attributes as height, weight, cephalic index, etc., have been found to be continuous, so that within the range of the scale used, any measure—integral or fractional—may exist and have meaning. Whenever gaps occur in a truly continuous series, therefore, these are usually to be attributed to our failure to measure enough cases, or to the relative crudity of our measuring instruments, or

to some other fact of the same sort, rather than to the fact that no measures exist within the gaps.

There are, however, measures which do not fall into continuous series. Thus a salary scale in a department store may run from $10 per week to $20 per week in units of 50 cents or $1.00; no one receives, let us say, $17.53 per week. Or, to take another example, the average family in a certain locality may work out mathematically to be 4.57 children, although there is obviously a real gap between four and five children. Series like these, which contain real gaps, are called discrete or discontinuous.

It is probably fortunate—at least from the standpoint of the beginner in statistics—that nearly all of the measures which we make in psychology are continuous or can be treated as continuous. This considerably simplifies the problem, inasmuch as we may concern ourselves (for the present at least) almost entirely with methods of handling continuous data, postponing the discussion of discrete series to a later page.

2. The Classification of Measures in Continuous Series

Data collected from test or experiment are often merely a series of numbers or mass of figures without meaning or significance until they have been rearranged or classified in some systematic way. The first task that confronts us, then, is the organization of our material, and this leads naturally to a grouping of the measures into classes or categories. The procedure in grouping falls under three main heads, which are given in order below:

(1) The determination of the *range:* the interval between the largest and the smallest measures. The range is easily found by subtracting the smallest from the largest measure.

(2) Deciding upon the number and size of the groups to be used in classification. The number and the size of these steps or class-intervals depend largely upon the range and the kind of measures with which we are dealing.

(3) The tabulation of the separate measures within their proper step- or class-intervals.

TABLE I

ARMY ALPHA SCORES MADE BY 54 COLUMBIA COLLEGE MEN

1. THE ORIGINAL SCORES (UNGROUPED)

185	174	127	183	168	*126	177	154	157	189	172
*201	158	160	179	184	155	137	177	164	198	176
188	197	151	188	188	169	195	165	185	188	164
195	176	185	185	179	146	182	153	158	160	191
176	138	185	155	178	151	144	191	170	157	

* Maximum score = 201. * Minimum score = 126.

2. THE SAME SCORES GROUPED INTO A FREQUENCY DISTRIBUTION BY THREE METHODS

(A) (1) Scores	(2) Tabulation	(3) F	(B) Scores	F	(C) Scores	F
200 up to 205	/	1	200–204.99	1	200–204	1
195 " " 200	////	4	195–199.99	4	195–199	4
190 " " 195	//	2	190–194.99	2	190–194	2
185 " " 190	~~////~~ ~~////~~	10	185–189.99	10	185–189	10
180 " " 185	///	3	180–184.99	3	180–184	3
175 " " 180	~~////~~ ///	8	175–179.99	8	175–179	8
170 " " 175	///	3	170–174.99	3	170–174	3
165 " " 170	///	3	165–169.99	3	165–169	3
160 " " 165	////	4	160–164.99	4	160–164	4
155 " " 160	~~////~~ /	6	155–159.99	6	155–159	6
150 " " 155	////	4	150–154.99	4	150–154	4
145 " " 150	/	1	145–149.99	1	145–149	1
140 " " 145	/	1	140–144.99	1	140–144	1
135 " " 140	//	2	135–139.99	2	135–139	2
130 " " 135		0	130–134.99	0	130–134	0
125 " " 130	//	2	125–129.99	2	125–129	2
		$N=54$		$N=54$		$N=54$

These three principles of classification are illustrated in Table I. The figures in this table represent the Army Alpha scores received by 54 college men. Since the highest score is 201, and the lowest 126, the range is found at once to be exactly 75 points. In deciding upon the number of "steps" or class-intervals to be used in grouping, the best general rule is to select by trial a step-interval which will yield not more than 20 nor less than 10 steps. The number of steps which a given interval will yield can be determined approximately (within one step)

by dividing the range by the step tentatively chosen. In the present problem, for example, 75 (the range) divided by 5 (the step-interval) gives 15, which is one less than the actual number of steps, namely 16. A step-interval of 3 points will yield approximately 25 steps, while a step-interval of 10 points will yield approximately 7.5 steps. (Actually, for the given data, a step-interval of 3 points yields 26 steps, and one of 10 points 8 steps.)

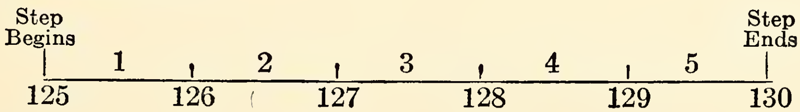
The tabulation of the separate scores within their appropriate step- or class-intervals is shown in Table I(2A). In the first column of this table,—in the column marked "Scores,"—the step-intervals have been listed serially, with the smallest measures at the bottom of the column. The first interval, "125 up to 130," begins at 125 and ends at 130; the second interval "130 up to 135," begins at 130 and ends at 135 and so on. The last interval, "200 up to 205," begins at 200 and ends at 205. In column 2, marked "Tabulation," the separate scores have been listed opposite their proper intervals. The first score, 185 [see Table I(1)], is represented by a tally placed opposite step-interval "185 up to 190"; the second score, 201, by a tally placed opposite step-interval "200 up to 205"; the third score, 188, by a tally placed opposite "185 up to 190" and so on for the other scores. When all 54 scores have been listed, the total number of tallies on each step-interval (i.e., the frequency) is written in column 3, headed F (frequencies). The sum of the F column is called N. In the present case, of course, $N = 54$. When the total frequency of each step-interval has been tabulated opposite its proper step-interval, as shown in column 3, our 54 Alpha scores are arranged into what is known as a Frequency Distribution.

The reader will note that the *lower limit* of the first step in the distribution (i.e., 125 up to 130) has been taken at 125 although the lowest actual score in the series is 126. This is due to the fact that when the step-interval equals 5 units, it facilitates tabulation as well as computations which come later on, if the lower limit of the first step-interval (and accordingly

of each succeeding step-interval) is a multiple of 5. A step-interval of 126 up to 131 is just as good as a step-interval of 125 up to 130, theoretically; the second, however, is much easier to handle from the standpoint of the arithmetic involved.

3. Three Ways of Expressing the Limits of a Step-interval

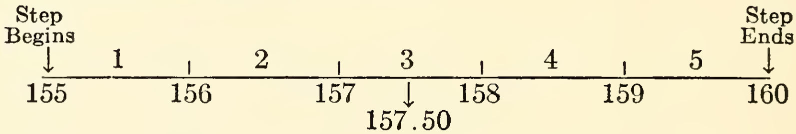
Table I (2 A,B,C) illustrates three ways of writing the limits of a step-interval. In (A) the interval "125 up to 130" means that all scores from 125 up to but not including 130 fall on this step. In (B) the step-interval 125–129.99 means exactly the same thing. The upper limit is written 129.99 simply to emphasize the fact that this step-interval includes score 129 *plus* fractional parts up to 130, but does *not* include score 130. (C) expresses the same facts more clearly than (A) and not so exactly as (B). Thus 125–129 means that this step-interval begins *with* score 125 and ends *with* score 129. A diagram will indicate how (A), (B), and (C) are simply three ways of expressing the same facts.



Either method (B) or method (C) is advised as preferable to (A). It is fairly easy—even when one is on guard—to let a score of say 160 slip into the step-interval 155 up to 160 due simply to the presence of the 160 at the upper limit of the step. The accurate tabulation of a frequency distribution depends on getting each score into its proper step-interval, and for this reason one cannot be too careful in defining the limits of the steps.

In any frequency distribution we always assume that the scores within a given interval (i.e., the frequency) are spread evenly over the entire interval; and this assumption holds whether the length of the step is 3, 5 or 10 units. If we wish to represent *all* of the scores within a given interval by some single value, however, the midpoint of the interval is taken as

the most logical choice. To illustrate, in the step-interval 155–159 [see Table I (2 C)] the six scores on this step are all represented by the same value, 157.50, the midpoint of the interval, although the scores are 155, 155, 157, 157, 158, 158. The reason why 157.50 is the midpoint of the step-interval can be shown graphically as follows:



A simple rule for finding the midpoint of a step is

$$\text{Midpoint} = \text{lower limit of step} + \frac{(\text{upper limit} - \text{lower limit})}{2}.$$

For example, in the present case, $155 + \frac{(160-155)}{2} = 157.50$.

Again, since the length of the step is 5, it follows that the midpoint must be 2.5 points from the lower limit of the step, i.e., at 155+2.5 or 157.50.

It is often a question whether the midpoint is a fair representative of all of the scores on a given step-interval. If we examine the six scores on step 155–159, two scores, the two 155's, are below the midpoint; two scores, the two 157's, are practically on the midpoint; and two scores, the two 158's, are above the midpoint. Also an examination of the step preceding and the step following 155–159 shows that on both of these steps there are 2 measures above and 2 below the midpoint. There seems good evidence, therefore, for assuming that the midpoint represents fairly the scores on these intervals, though it is true that the balancing of scores above and below the midpoint is not always as clear cut as in the examples cited. In certain cases, in fact (e.g., when the distribution is considerably "skewed" [1]), there are often many more scores on one side of the midpoint than the other, and the midpoint assumption is

[1] When the scores are "piled" up at either the lower or the upper end of the scale, the distribution is said to be "skewed." See page 86.

then clearly untenable. The fact remains, however, that in most frequency distributions of mental and educational measurements, especially when the number of measures is large, the assumption that the midpoint represents all of the scores on the interval is a valid one, since in the long run about as many scores will fall above as below the midpoint value.

4. The Meaning of a Single Score in a Continuous Series

So far we have discussed the classification of scores into step-intervals (the frequency distribution) and the necessity of defining carefully the upper and lower limits of our step-intervals. We shall now try to give a more precise notion of what is meant by a single score, for example, a score of 165 points on Army Alpha. If we think of the score 165 as occupying a certain interval or distance on a linear scale, then any fractional value from 165 up to (but not including) 166, e.g., 165.3, 165.8, etc., will fall within this interval and be scored simply as 165. See illustration:

STEP 165

165 166

A score of 165 may mean, therefore, that the person who made it was just barely through 165 items, or that he had nearly completed 166—in either case his score will be 165.

In *performance* scales a score equal to or greater than 8, say, but less than 9 is placed on step 8–9 or 8–8.99 and scored 8. In most *product* scales, however,—the Thorndike Handwriting Scale is an example—a score of 8 represents any value from 7.5 to 8.5: i.e., any value from a point one half step below 8 to a point one half step above. Thus scores 7.7, 8.0, 8.4, etc., would all be scored 8. If as before we think of a score on such a scale as a linear magnitude, 8 represents the midpoint of that interval which extends from 7.5 to 8.5. See illustration:

STEP 8

7.5 8 8.5

This method of scoring is employed in scales which measure handwriting, drawing, composition, etc.

It is evident from the foregoing that the meaning of a single score in a continuous series will depend upon how the test is scored. If the score is not defined by the test, it is probably safer to assume that a score of 22, say, means 22–23, rather than 21.5–22.5.

II. Measures of Central Tendency

When scores or other measures have been tabulated into a frequency distribution, generally our next task is to find a measure of central tendency. The value of a measure of central tendency is twofold: in the first place, it is a single measure which represents *all* of the scores made by the group, and as such gives a concise description of the performance of the group as a whole; secondly, it enables us to compare two or more groups in terms of typical performance. There are three measures of central tendency in common use, (1) the average or arithmetic mean, (2) the median, and (3) the mode. We shall consider these three measures in order.

1. The Average, or Arithmetic Mean [1]

The average is the best known of the measures of central tendency. It may be defined simply as the sum of the separate scores or measures in a series divided by their number. To illustrate, if a man makes \$3.00, \$4.00, \$3.50, \$5.00 and \$4.50 on five successive days, his average daily wage (\$4.00) is obtained by dividing the sum of his daily earnings by the number of days he has worked. The formula for the average of a series of ungrouped measures is simply

$$\text{Average} = \frac{\Sigma(\text{Measures})}{N}, \qquad (1)$$

in which N is the number of measures in the series.[2]

[1] The term " average " is often used as a general expression to cover any measure of central tendency. It is here used in a more restricted sense.

[2] The symbol Σ means "sum of."

When measures have been grouped into a frequency distribution, it is necessary to calculate the average by a slightly different method from the one given above. The two illustrations in Table II will make this method clear. The first of these shows the calculation of the average for the 54 Army Alpha scores which we have already tabulated into a frequency distribution in Table I. Note that we first calculate the $F \times M$ column by multiplying the midpoint (M) of each step-interval by the number of scores (F) on it; and that the average (171.57) is then simply the sum of the $F \times M$ (9265) divided by N (54). The use of the midpoint for all of the scores on the interval is made necessary by the fact that when scores have been grouped into step-intervals they lose their identity and are thereafter represented by the midpoint of the particular interval on which they happen to fall. Hence, we must multiply or "weight" the midpoint of each step (M) by the frequency (F) on that step; add the $F \times M$, and divide by N to get the average. The formula may be written

$$\text{Average} = \frac{\Sigma(F \times M)}{N}. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Example (2), Table II, is a second illustration of the calculation of an average from grouped data. This frequency distribution represents 200 scores made by a group of adults on a cancellation test. These scores are classified into 9 steps; and since the step-interval is 4 points, the midpoint of each step is found by adding $\frac{1}{2}$ of 4 to the beginning of each step (for example, $104+2=106$). The $F \times M$ column (found as shown above) totals 23988, and N equals 200. Hence, applying formula (2), the average is found to be 119.94.

In both illustrations in Table II we have found the average of the scores made by a given *group*. There is no reason, however, why we cannot use either formula (1) or (2) to find the average of a number of measurements made on the *same individual*, as well. Thus an individual's reaction time to light may be measured 100 times, the measures tabulated into a

TABLE II

To Illustrate the Calculation of the Average, Median, and Mode, from Data Grouped into a Frequency Distribution

1. DATA FROM TABLE I (2), 54 ARMY ALPHA SCORES
THE STEP-INTERVAL = 5 POINTS

Scores	Midpoint	F	$F \times M$
200–204.99	202.5	1	202.50
195–199.99	197.5	4	790.00
190–194.99	192.5	2	385.00
185–189.99	187.5	10	1875.00
180–184.99	182.5	3	547.50
175–179.99	177.5	8	1420.00
		—26	
170–174.99	172.5	3	517.50
165–169.99	167.5	3	502.50
160–164.99	162.5	4	650.00
155–159.99	157.5	6	945.00
150–154.99	152.5	4	610.00
145–149.99	147.5	1	147.50
140–144.99	142.5	1	142.50
135–139.99	137.5	2	275.00
130–134.99	132.5	0	......
125–129.99	127.5	2	255.00
		$N=54$	9265.00

(1) Average $= \dfrac{\Sigma(F \times M)}{N} = \dfrac{9265}{54} = 171.57.$

(2) $\left(\dfrac{N}{2} = 27\right)$ Median $= 175 + \dfrac{1}{8} \times 5 = 175.625.$

(3) Crude mode falls on class-interval, 185–189.99 or at 187.5

2. SCORES MADE BY 200 ADULTS ON A CANCELLATION TEST
STEP-INTERVAL = 4 POINTS

Scores	Midpoint	F	$F \times M$
136–139	138	3	414
132–135	134	5	670
128–131	130	16	2080
124–127	126	23	2898
120–123	122	52	6344
116–119	118	49	5782
		—52	
112–115	114	27	3078
108–111	110	18	1980
104–107	106	7	742
		$N=200$	23988

(1) Average $= \dfrac{\Sigma(F \times M)}{N} = \dfrac{23988}{200} = 119.94.$

(2) $\left(\dfrac{N}{2} = 100\right)$ Median $= 116 + \dfrac{48}{49} \times 4 = 119.92.$

(3) Crude mode falls on class-interval, 120–123, or at 122.

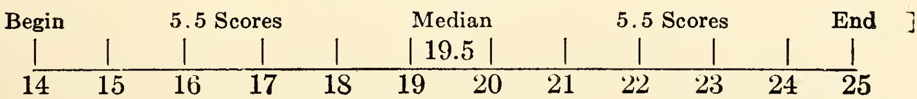
frequency distribution, and the average found in exactly the same way in which we find the average reaction time to light of 100 different observers.

2. The Median

When scores or other measures are arranged in order of size, the median is defined as the midpoint of the series, that is, as the point above which and below which are 50% of the measures. By definition, therefore, the median may be found by counting off one half of the measures, i.e., $\frac{N}{2}$, from either end of the series.

Let us first consider the calculation of the median for scores or measures in a simple ungrouped series. Two cases arise: Case I when N is odd, and Case II when N is even. As an illustration of the first case, take the following eleven consecutive scores: 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24. Now since N equals 11, $\frac{N}{2} = 5.5$; and counting off the first five scores, namely, 14, 15, 16, 17, 18, we reach 19, since score 18 means "18 up to 19." (See page 7.) The .5 left of our 5.5 then locates the median midway between 19 and 20, viz., at 19.5. To verify this result we may count off 5.5 scores beginning at the other end of the series. The five scores, 24, 23, 22, 21, and 20, take us to 20 (the upper limit of score 19) and the .5 left puts the median at a point midway on the scale between 20 and 19, viz., at 19.5 again. (See diagram below.)

Case I (N is odd)



To illustrate the procedure when N is even, let us drop off the first score (14) from the series of eleven scores in Case I. N is now 10, and $\frac{N}{2}$ is 5.0. Counting off the first five scores, therefore,

from the *small* end of the series, i.e., 15, 16, 17, 18, 19, we reach 20 (the *upper* limit of "19 up to 20") as the median. Likewise, if we count down five scores from 24, i.e., 24, 23, 22, 21, 20, we again reach 20, the *lower* limit of the step "20 up to 21." See diagram below:

Case II (N is even)

Begin		(5 Scores)			Median		(5 Scores)			End
15	16	17	18	19	20	21	22	23	24	25

It will be noted that in the two cases just cited, the measures were taken to be in continuous series. If, instead of continuous, the eleven scores under Case I are taken as *discrete* or discontinuous there is now no value which fulfills the definition of the median as the *midpoint* in the series. When N is odd, however, the *midscore* or the *middle measure* may be obtained by counting off $\frac{(N+1)}{2}$ scores from either end of the series, after the scores have been arranged in order of size from least to greatest. Thus, (Case I) $\frac{11+1}{2}$ or 6 scores counted off from either end of our series puts the midscore at 19—since there are 5 scores above and 5 scores below this score. A slightly different procedure is necessary when N is even. If the ten scores under Case II, for example, are taken as discrete, there is in this series, clearly no median value, and no midscore. However, in such cases as this it is customary to take the midscore arbitrarily at a point *midway* between the two middlemost scores. Thus, in our illustration, $\frac{N+1}{2}=5.5$, which puts the midscore at 19.5, midway between 19 and 20, the two middlemost scores. (For a discussion of the median for discrete measures grouped into a frequency distribution, see page 36.)

The method of calculating the median for continuous data grouped into a frequency distribution is shown in the two examples in Table II. Since there are 54 scores in the first

distribution, $\frac{N}{2}$ is 27. The median, therefore, is that point on the scale which has 27 scores on each side of it. If we begin at the small end of the distribution[1] and add up the scores in order, the step-intervals 125–129.99 to 170–174.99, inclusive, are found to contain just 26 scores. The next step, 175–179.99, contains 8 scores (assumed to be evenly spread over the entire step. See page 5.) To get the 1 extra score needed to make 27, therefore, we must take $1/8 \times 5$—the length of step—and add this amount (.625) to 175, the beginning of the step-interval 175–179.99. This puts the midpoint at $175 + .625$ or 175.625, which is, accordingly, the median of the distribution. (See Diagram I.)

A second illustration of how the median is found when the data are grouped into a frequency distribution is given in Table II (2). This second example should aid in clearing up any doubtful points in the first problem. Since there are 200 scores in this distribution, one half of the scores is 100, and the median must lie at a point 100 scores distant from either end of the distribution. If we begin at the small end of the distribution, i.e., at 104–107, and add the scores in order, 52 scores will take us *through* step 112–115. The 49 scores on the next step-interval, (116–119) total 101 scores—one too many to give us the median. To get the 48 scores needed to make exactly 100, therefore, we must take $48/49 \times 4$ (the length of the step) and add this amount, 3.92 to 116, the beginning of the step-interval. This takes us exactly 100 scores into the distribution, and locates the median at 119.92. Diagram I (2) shows graphically how this median is obtained.

Summary of the steps in computing the median from data tabulated in a frequency distribution:

(1) Find $\frac{N}{2}$ measures.

[1] While the median may be found equally well by counting in $\frac{N}{2}$ scores from the large end of the distribution, it is simpler to begin at the small end, and the student is advised to follow this plan first.

(2) Begin at the smaller end of the distribution and count the measures serially up to the interval which contains the median.

(3) Divide the number of measures necessary to fill out $\frac{N}{2}$ by the frequency on the interval containing the median [reached

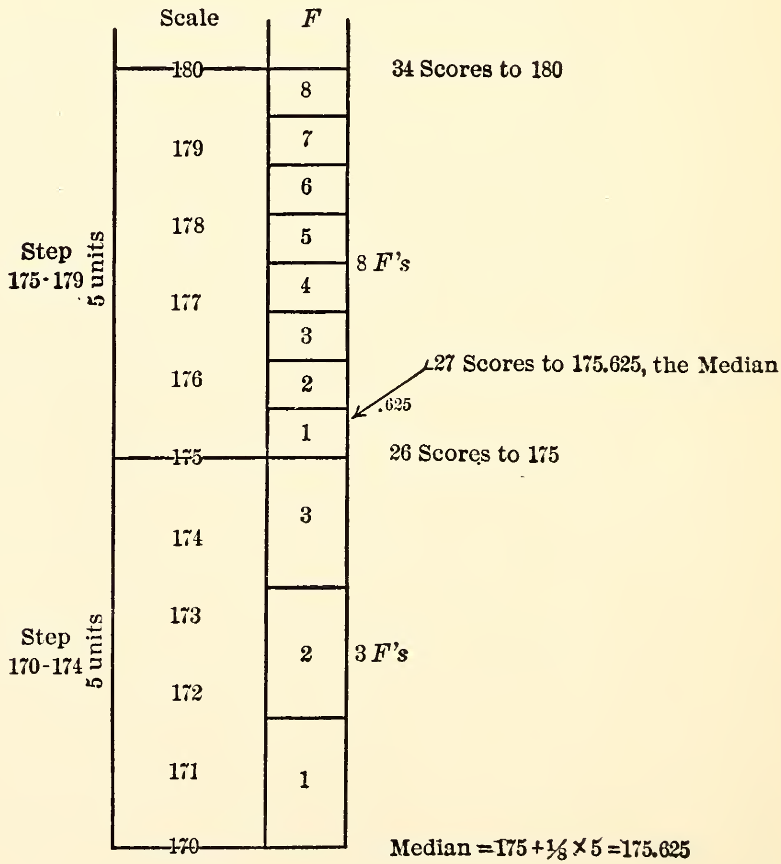


DIAGRAM I (1)

THE CALCULATION OF THE MEDIAN.

EXPLANATION—26 scores go up to 175 on the scale; 34 scores to 180. To find how far 27 scores will go, we must take $\frac{1}{8}$ of 5 (the step length) and add this to 175. This puts the median at 175.625.

in (2) above] and multiply the result by the length of the step-interval.

(4) Add the amount obtained in (3) to the lower limit of

the step which contains the median. This will give the median point on the scale.

3. The Mode

The mode is most simply defined as that measure which occurs most often in a series. In the series, 10, 11, 11, 12, 12,

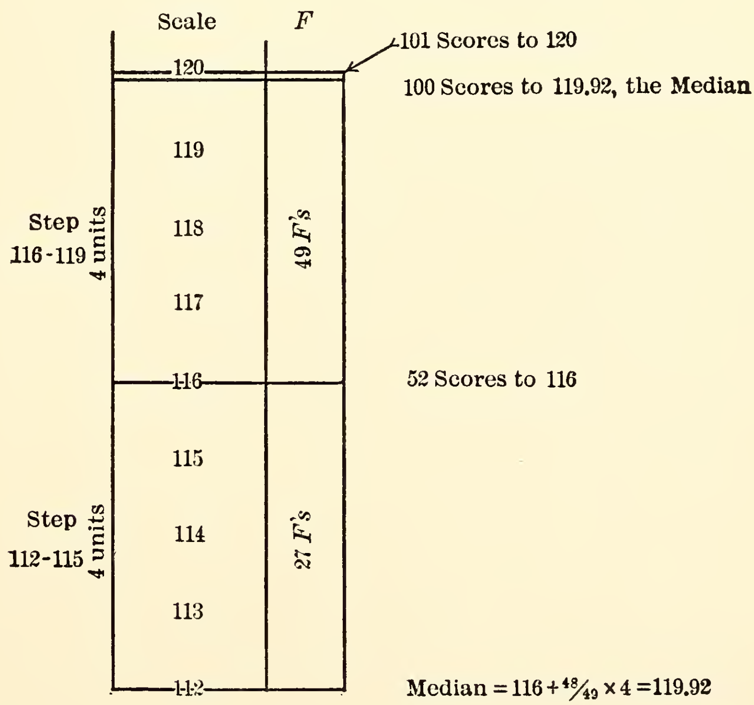


DIAGRAM I (2)

THE CALCULATION OF THE MEDIAN.

EXPLANATION—52 scores counted off take us to 116 on the scale; 101 scores take us to 120. To find how far 100 scores go, we must take 48/49 of 4 (the step length) and add this amount (3.92) to 116. This locates the median at 119.92.

13, 13, 13, 14, 14, and 15, for example, since the most often recurring measure is 13 this measure may be taken as the mode. In Table I (1) we find from the ungrouped scores that 185 occurs 5 times—more often than any other single score—and hence 185 may be taken as the mode of this series.

When the scores or measures are continuous and have been grouped into a frequency distribution, the "crude mode" is often taken as the midpoint of the step-interval which contains the greatest frequency. In Table I, for example, if we did not know from the ungrouped scores that 185 is the modal score, the crude mode of the distributions given in (2) would be taken at 187.50, the midpoint of step 185–189, the step-interval containing the greatest frequency. Likewise, in Table II, the crude mode would be 122, the midpoint of the step which contains the greatest frequency.

It is clear that the crude mode will be dependent to a large extent upon the size of the step-interval selected (i.e., on whether the grouping is by large or small steps) and for this reason it is often an unstable measure of central tendency. This is not necessarily a serious drawback, however, as the mode is usually employed simply to indicate in a rough way the center of concentration in the distribution. For this purpose it is not necessary to define it so carefully as we do the median or the arithmetic mean.

III. Measures of Variability

In Section II we discussed the calculation of the so-called "measures of central tendency"—measures typical or representative of the set of scores as a whole. Our next step is the calculation of the *variability* of the scores, i.e., of the "scatter" or "spread" of the separate scores or measures around their measure of central tendency. This will be the task of the present section.

The usefulness of some measure of variability can be shown by a simple example. Suppose that we have given a test of controlled association to a group of 50 boys and the same test to a group of 50 girls. The average scores are, Boys, 34.6 secs., and Girls, 34.5 secs.—so far as the averages go, there is apparently no difference in the performance of the two groups. Suppose, however, that on examining the original scores, we

find the boys' scores ranging from 15 to 51 secs. and the girls' scores ranging from 19 to 45 secs. This discovery would make it evident at once that in a general way, the boys "cover more territory"—are more variable—than the girls, and this greater variability may be of considerably more interest than the lack of difference in the average scores. If a group is homogeneous, i.e., made up of individuals of nearly the same ability, most of the scores will fall near the same point on the scale, the range will be relatively short, and the variability will be small. If, however, the group contains individuals of widely differing capacity, the scores will be strung out from high to low, the range will be relatively wide, and the variability will be large. Four measures have been devised to take account of this factor of variability within a set of measures. These are (1) the range, (2) the quartile deviation, or Q, (3) the average deviation, or AD, and (4) the standard deviation, or SD.

1. The Range

In grouping the scores in Table I into a frequency distribution (page 3) we have already had occasion to use the range. It may be re-defined simply as the interval between the largest and the smallest measures. In the illustration given above, the range of the boys' scores is 51–15 or 36, and the range of the girls' scores 45–19 or 26. The range is the most general measure of "spread" or "scatter." It includes 100% of the distribution, and is employed when we wish to make a rough comparison of two or more groups for variability; or when the number of measures is too small to justify the calculation of some more refined measure of variability. Since the range only takes account of the extremes of the series, it is obviously unreliable when frequent or large gaps occur in the distribution of scores.

2. The Quartile Deviation, or Q

The quartile deviation, or Q, may be defined as one half of the distance between the 75th and the 25th percentile points in the given distribution. The 25th percentile, or Q_1, is the

first quarter or quartile point on the scale; the point below which lie 25% of the measures. In like manner, the 75th percentile, or Q_3, is the third quarter or quartile point on the scale, the point below which lie 75% of the measures. (By analogy, the median is Q_2, the second quartile point.)

In order to find Q, it is obvious that we must first calculate the 75th and 25th percentile points. These points are found in exactly the same way as the median: viz., to find Q_1 we count off 25% of the scores from the beginning of the distribution; and to find Q_3, we count off 75% of the scores from the beginning of the distribution.

Table III illustrates the calculation of Q for the distribution of 54 Alpha scores tabulated in Table I. First, to find Q_1, we must count off 1/4 of the total number of scores, i.e., 13.5, from the small end of the distribution. When the scores (the F's) are added in order the first six step-intervals (the steps 125–129.99 to 150–154.99 inclusive) are found to contain 10 scores. The next step, 155–159.99, contains 6 scores.[1] We need only 3.5 additional scores, however, to make up the necessary 13.5; hence we take $\frac{3.5}{6}\times 5$ (the step length) and add this amount (2.92) to 155, the beginning of the step. This locates Q_1 at $155+2.92$ or 157.92.

In like manner, we find Q_3 by counting off 3/4 of the scores from the small end of the distribution. 3/4 of $N=40.5$; and the F's on steps 125–129.99 to 180–184.99, inclusive, added in order, total 37. The next step, 185–189.99, contains 10 scores. To round out the necessary 40.5, therefore, we take $\frac{3.5}{10}\times 5$ (the step length) and add this amount (1.75) to 185, the beginning of the step. This puts Q_3 at 186.75 since 40.5 scores reach this point.

[1] Assumed to be spread evenly over the entire step. See page 5.

TABLE III

To Illustrate the Calculation of Q, AD, and SD from Data Grouped into a Frequency Distribution

1. DATA FROM TABLE I, 54 ARMY ALPHA SCORES

midpoint — average

(1) Scores	(2) Midpoint	(3) F	(4) D	(5) FD	(6) FD^2
200–204.99	202.50	1	30.93	30.93	956.66
195–199.99	197.50	4	25.93	103.72	2689.46
190–194.99	192.50	2	20.93	41.86	876.13
185–189.99	187.50	10	15.93	159.30	2537.65
180–184.99	182.50	3	10.93	32.79	358.39
175–179.99	177.50	8	5.93	47.44	281.32
170–174.99	172.50	3	.93	2.79	2.79
165–169.99	167.50	3	− 4.07	−12.21	49.69
160–164.99	162.50	4	− 9.07	−36.28	329.06
155–159.99	157.50	6	−14.07	−84.42	1187.79
150–154.99	152.50	4	−19.07	−76.28	1454.66
145–149.99	147.50	1	−24.07	−24.07	579.36
140–144.99	142.50	1	−29.07	−29.07	845.06
135–139.99	137.50	2	−34.07	−68.14	2321.53
130–134.99	132.50	0	−39.07	....	....
125–129.99	127.50	2	−44.07	−88.14	3884.33
		$N=54$		837.44	18353.88

Average = 171.57 (Table II)

$$\frac{N}{4}=13.5, \text{ therefore,} \qquad \frac{3N}{4}=40.5, \text{ therefore,}$$

$$Q_1=155+\frac{3.5}{6}\times 5=157.92 \qquad Q_3=185+\frac{3.5}{10}\times 5=186.75$$

$$Q=\frac{Q_3-Q_1}{2}=\frac{186.75-157.92}{2}=14.42$$

$$AD=\frac{\Sigma FD}{N}=\frac{837.44}{54}=15.51$$

$$SD=\sqrt{\frac{\Sigma FD^2}{N}}=\sqrt{\frac{18353.88}{54}}=\sqrt{339.887}=18.44$$

TABLE III—Continued

2. DATA FROM TABLE II(2), 200 CANCELLATION SCORES

(1) Scores	(2) Midpoint	(3) F	(4) D	(5) FD	(6) FD^2
136–139	138	3	18.06	54.18	978.49
132–135	134	5	14.06	70.30	988.42
128–131	130	16	10.06	160.96	1619.26
124–127	126	23	6.06	139.38	844.64
120–123	122	52	2.06	107.12	220.67
116–119	118	49	− 1.94	− 95 06	184.42
112–115	114	27	− 5.94	−160.38	952.66
108–111	110	18	− 9.94	−178.92	1778.47
104–107	106	7	−13.94	− 97.58	1360.27
		$N=200$		1063.88	8927.30

Average = 119.94 (Table II)

$\frac{N}{4}=50$, therefore,

$$Q_1=112+\frac{25}{27}\times 4=115.70$$

$\frac{3N}{4}=150$, therefore,

$$Q_3=120+\frac{49}{52}\times 4=123.77$$

$$Q=\frac{Q_3-Q_1}{2}=\frac{123.77-115.70}{2}=4.04$$

$$AD=\frac{\Sigma FD}{N}=\frac{1063.88}{200}=5.32$$

$$SD=\sqrt{\frac{\Sigma FD^2}{N}}=\sqrt{\frac{8927.30}{200}}=6.68$$

With Q_1 and Q_3 known, the quartile deviation, Q, is easily calculated from the formula

$$Q=\frac{Q_3-Q_1}{2} \qquad (3)$$

In the present problem, $Q=\frac{186.75-157.92}{2}$ or 14.42.

A second illustration of the calculation of Q from a frequency distribution is given in Table III (2). Since the N of this dis-

tribution is 200, 1/4 of the measures equals 50. The steps 104–107 and 108–111 contain 25 scores; and the next step contains 27 scores. To find the point reached by 50 scores, therefore, we must take $25/27 \times 4$ (the step length) and add this amount (3.70) to 112, the lower limit of step 112–115. This locates Q_1 at 115.70.

To find Q_3, we must count off 3/4 of N or 150 scores from the small end of the distribution. The first four steps include 101 scores, and the next step, 120–123, contains 52. To fill out 150, therefore, we take $49/52 \times 4$ (the length of step) and add this increment (3.77) to 120 to locate Q_3 at 123.77. Substituting 115.70 for Q_1 and 123.77 for Q_3 in formula (3) we get a Q of 4.04 points.

The quartile points, Q_1 and Q_3, are of considerable importance in that they mark off the limits within which fall the middle 50% of the measures in the distribution. The distance between these two points is often called the *interquartile range;* hence Q is sometimes called the Semi-interquartile Range. Q actually measures the average distance of the two quartile points from the median, and because of the ease with which it can be found is a valuable measure of the closeness with which the scores are grouped directly around the median point. If the scores of a distribution are closely packed together, the quartiles will be close together and Q will be small; if the scores are scattered, the quartiles will be relatively far apart, and Q will be large.

When the distribution is symmetrical or " normal " (see page 85) Q marks off exactly the limits of the 25% of the cases just above, and the 25% of the cases just below the median: and accordingly, the median lies just halfway between the two quartile points Q_1 and Q_3. Q is then commonly known as the *PE* (probable error). The terms Q and *PE* are often used interchangeably, although it is probably best to restrict the use of the latter term to normal distributions, and to the measurement of reliability. The value of the *PE* as a measure of reliability will be discussed at length in Chapter III.

Summary of Steps in Calculation of Q (Data Grouped)

To find Q_1:

1. Divide N by 4.
2. Begin at the small end of the distribution, and count the scores up to the interval which contains Q_1.
3. Divide the number of measures necessary to locate Q_1 $\left(\text{i.e., to complete } \frac{N}{4}\right)$ by the frequency in the interval reached in (2) above, and multiply the result by the step-interval.
4. Add the amount obtained in (3) to the lower limit of the step-interval on which Q_1 lies. The result is Q_1.

To find Q_3:

1. Find 3/4 of N.
2. Begin as before at the small end of the distribution, and count up the scores until the interval which contains Q_3 is reached.
3. Divide the number of scores required to locate Q_3 by the frequency in the interval reached in (2) and multiply the result by the step-interval.
4. Add the amount obtained in (3) to the lower limit of the step-interval on which Q_3 lies. This locates Q_3.

To find Q:

Substitute Q_3 and Q_1 in formula (3),

$$Q = \frac{Q_3 - Q_1}{2}.$$

3. The Average Deviation, or AD

The average deviation or AD (also written mean deviation or MD) may be defined as the average of the deviations of all the separate measures in a series taken from their central tendency (usually the average, less frequently the median,

or mode). In averaging deviations to find the *AD*, no account is taken of signs, and all deviations, whether positive or negative, are treated as positive.

An example will make the definition clearer. If we have five scores, 6, 8, 10, 12, and 14, the average is easily found to be 10. It is then a simple process also to find the deviation of each measure from the average by subtracting the average from each measure. Thus 6, the first score, minus 10 equals -4 (calculation algebraic); $8-10=-2$; $10-10=0$; $12-10=2$; and $14-10=4$. The five deviations measured from the average are -4, -2, 0, 2, and 4. Now adding these deviations without regard to sign, the sum is 12; and dividing 12 by 5, we get 2.4, as the *average* of the 5 deviations from the average, or the *AD*. The formula for the *AD* with simple ungrouped numbers like these may be written,

$$AD=\frac{\Sigma D}{N}\ \text{(arithmetical)}, \quad . \quad . \quad . \quad . \quad . \quad (4)$$

in which $\Sigma D=$ sum of deviations, and N is, as before, the number of cases or items in the series.

In Table III, the calculation of the *AD* for scores grouped into a frequency distribution is illustrated by two problems. The average of problem (1) has already been found in Table II to be 171.57. Hence, to find the average deviation of the scores in this distribution from the average, we must take our deviations (*D*'s) around this point. Note, however, that, since the scores have been grouped into step-intervals, we are no longer able to get the *D* of *each* score from the average; and hence we simply find the deviation (*D*) of the midpoint of *each step* from the average. The substitution of the midpoint value for *all* of the scores within the step is the only difference between the computation of *D*'s with grouped and ungrouped measures. For example the *D* of step 200-204.99 is 30.93, found by subtracting 171.57 (the average) from 202.50 (the midpoint of the step). Likewise, the *D* of the next step is 25.93, found by subtracting 171.57 from 197.50. All of the *D*'s

are positive as far down the scale as 170–174.99, as in each case the midpoint is larger numerically than the average. From the step-interval 165–169.99 on down to the beginning of the series, however, the D's are negative, as the midpoints of these steps are all smaller than 171.57. Thus the D of step 165–169.99 is -4.07, e.g., $167.50-171.57=-4.07$; and the D of the lowest step in the distribution, 125–129.99, is -44.07.

It will be helpful in finding deviations to remember that the average is *always* subtracted from the individual score or midpoint value. That is,

Deviation = Score or Midpoint − Average (calculation algebraic).

Hence it is clear that when the score or midpoint is numerically *larger* than the average, the deviation must be positive; when the score or midpoint is numerically *smaller* than the average, the deviation must be negative.

It is obviously unnecessary to subtract the average from each midpoint *separately* in order to obtain the different D's. The reason, of course, is that each step-interval is 5 points; hence, after finding the D of step 200–204.99 to be 30.93, we need only subtract 5 points from this D in order to obtain 25.93, the D of the next step; then 5 again to obtain 20.93, the D of the next step, and so on.[1] The negative D's are obtained in exactly the same way as the positive D's. Thus $.93-5=-4.07$; $-4.07-5=-9.07$ and so on to -44.07.

Column 4 gives the deviation of *each step-interval* (as represented by its midpoint) from the average of the distribution. There are, however, more scores on some steps than on others; and for this reason each midpoint-deviation (D) in column 4 must be "weighted" (multiplied) by the number of scores (F) which it represents. This gives the FD column,—column 5. The first FD is 30.93; for since there is only 1 score on step 200–204.99, we need simply multiply the first D by 1. The next FD is 103.72; since each

[1] Checking the D's occasionally to avoid carrying an error throughout our calculations.

of the 4 scores on step 195–199.99 has a D of 25.93. In like manner, we obtain the other FD's, by multiplying each D in column 4 by its corresponding frequency (F) in column 3.

When all of the FD's have been calculated, we sum the column without regard to sign and divide by N to obtain the AD. In the present problem, the AD equals $\frac{837.44}{54}$ or 15.51.

The formula for the AD for measures grouped into a frequency distribution may now be written as follows:

$$AD = \frac{\Sigma(F \times D)}{N} \text{ (arithmetical)}. \quad . \quad . \quad . \quad . \quad (5)$$

This formula applies equally well to the AD found from the average, median, or mode.

The second problem in Table III shows the calculation of the AD for the 200 cancellation scores, grouped into a frequency distribution with a step of 4. The average for this distribution has been found to be 119.94 (see Table II, 2). Hence, the D of the first step 136–139 (midpoint 138), from the average is 18.06. The next D may be found by subtracting 4 (the step-interval) from 18.06, and each succeeding D in turn by subtracting 4 from the D just preceding it.

The FD's in column 5 are found [as previously shown in (1)] by "weighting" each D by the F which it represents,—by the F opposite it. The sum of the FD column is 1063.88; and since N is 200, from formula (5) we obtain 5.32, as the AD of the scores in this distribution from their average 119.94.

In a perfectly symmetrical or normal distribution (page 85) the AD—when measured off above and below the average—marks the limits of the middle 57.5% of the measures. Thus the AD is seen to be slightly larger than the Q. In general, a large AD means that the scores in the distribution are scattered around the central tendency; a small AD means that they are concentrated within a relatively narrow range.

4. The Standard Deviation, or *SD*

The standard deviation or *SD* is the most reliable of the measures of variability, and for this reason is customarily used in research which requires great accuracy. The *SD* differs from the *AD* in several respects. In the first place, in calculating the *AD* we disregard signs and treat all deviations as positive; in finding the *SD*, on the other hand, we avoid this difficulty of signs by squaring the separate deviations. Again, the deviations used in computing the *SD* are always taken from the average, and never from the median or mode as is sometimes done in finding an *AD*. The conventional symbol used to denote the *SD* is the Greek letter sigma, σ.

We may define the *SD* or σ as the square root of the mean (or average) of the squared deviations taken from the average of the distribution. To illustrate the calculation of the *SD* in a simple case, let us consider the example used to illustrate the calculation of the *AD* (see page 25) in which the deviations of the five measures, 6, 8, 10, 12, and 14, from their average 10 were found to be -4, -2, 0, 2, and 4, respectively. If we square each of these deviations we get 16, 4, 0, 4, and 16 (the minus signs become plus in squaring). Next, summing up these five squares and dividing by 5, the mean of the squares (8) is obtained; extracting the square root of this result gives 2.828 the *SD* or σ of the series. The formula for the σ of a series of numbers, ungrouped, is

$$\sigma=\sqrt{\frac{\Sigma D^2}{N}}. \qquad (6)$$

Table III illustrates the calculation of σ for scores grouped into a frequency distribution. The process is identical with that used for simple numbers except that in addition to squaring the D of each midpoint from the average, we "weight" each of these squared deviations by the frequency which it represents—the frequency opposite it. This gives the FD^2 column. By simple algebra, $D \times FD = FD^2$; and accordingly the easiest way to obtain the entries in this column is by

multiplying the corresponding D's and FD's in columns 4 and 5. The first FD entry, for example, is 956.66, the product of 30.93×30.93; the second is 2689.66, the product of 103.72×25.93, and so on to the end of the column. All of the FD^2's are necessarily positive, since each negative D is matched by a negative FD and consequently the product is positive. The sum of the FD^2 column (18,353.88) divided by $N(54)$ gives the mean of the squared deviations as 339.887; and the square root of this result is 18.44, the standard deviation. The formula for the SD when the data are grouped into a frequency distribution is

$$\sigma = \sqrt{\frac{\Sigma FD^2}{N}}. \quad \cdots \cdots \quad (7)$$

Problem (2) of Table III furnishes another illustration of the calculation of σ from grouped data. Column 6, the FD^2 column has been obtained, as in the previous problem, by multiplying each D by its corresponding FD. The sum of the FD^2 column is 8927.30; and N is 200. Hence, applying formula (7) we get 6.68 as the standard deviation [see Table III, (2) for calculations].

The standard deviation is, in general, less affected by chance fluctuations than the AD, and is, therefore, a more stable measure of dispersion. In a "normal" distribution (page 85) the SD when measured off above and below the average marks the limits of the middle 68.26% (roughly the middle 2/3) of the distribution. This is approximately true, also, for less symmetrical distributions. For example, in the first problem in Table III, the middle two thirds of the scores will fall roughly between score 190 (171.57+18.44) and score 153 (171.57−18.44). The standard deviation is always larger than the AD which, in turn, is always larger than Q. This relation supplies a rough but simple check on the accuracy of calculated measures of variability.

IV. The Short Method of Finding the Average, the *AD*, and the $SD(\sigma)$

In Tables II and III, the average, the *AD*, and the *SD* have been calculated by what is oftentimes known as the Long Method. The reader will recall that the average in these tables was found by multiplying the midpoint of each step-interval by the number of scores on the step, summing up this column (the $F \times M$) and dividing by N, the number of cases (page 9). Besides, in finding the *AD* and the *SD* all midpoint deviations were figured from the actual averages of the distributions.

It is, no doubt, already apparent that the Long Method (*LM*) requires the handling of large numbers and decimals and that the calculations are often tedious. To save time and labor, therefore, the Guessed Average Method, or more simply the Short Method (*SM*), has been devised for the express purpose of cutting down the calculations involved in finding the average, the *AD*, and the *SD*. (The Short Method does not apply to the computation of the Median and the *Q*, which are always found by the methods with which we are already familiar.) The student of statistics should make a special effort to learn the Short Method to the point where he can use it with facility. Not only is it a great time and labor saver, but in the calculation of coefficients of correlation it is well-nigh indispensable.

Table IV (2) illustrates the calculation of the average, *AD*, and *SD* by the Short Method. In order to make a comparison of the computations involved in the two methods easier, the calculations by the Long Method of the average, *AD*, and *SD* for the same data are also given in the Table.

1. The Calculation of the Average by the Short Method

The first important fact to grasp in beginning a study of the calculation of the average by the Short Method is that we "guess" or assume an average at the outset, and later apply

TABLE IV

To Illustrate the Calculation of the Average, AD, and SD by the Short Method. Data from Table II (1) Calculations for Long Method Given for Comparison.

1. LONG METHOD

(1) Scores	(2) Midpoint	(3) F	(4) $F \times M$	(5) D	(6) FD	(7) FD^2
200–204	202.5	1	202.5	30.93	30.93	956.66
195–199	197.5	4	790.0	25.93	103.93	2689.46
190–194	192.5	2	385.0	20.93	41.86	876.13
185–189	187.5	10	1875.0	15.93	159.30	2537.65
180–184	182.5	3	547.5	10.93	32.79	358.39
175–179	177.5	8	1420.0	5.93	47.44	281.32
170–174	172.5	3	517.5	.93	2.79	2.59
165–169	167.5	3	502.5	− 4.07	−12.21	49.69
160–164	162.5	4	650.0	− 9.07	−36.28	329.06
155–159	157.5	6	945.0	−14.07	−84.42	1187.79
150–154	152.5	4	610.0	−19.07	−76.28	1454.66
145–149	147.5	1	147.5	−24.07	−24.07	579.36
140–144	142.5	1	142.5	−29.07	−29.07	845.06
135–139	137.5	2	275.0	−34.07	−68.14	2321.53
130–134	132.5	0	.....	....	....	....
125–129	127.5	2	255.0	−44.07	−88.14	3884.33
		$N=54$	9265.0		837.44	18353.88

1. $\text{Aver.}=\dfrac{\Sigma FM}{N}=\dfrac{9265}{54}=171.57$

2. $AD=\dfrac{\Sigma FD}{N}=\dfrac{837.44}{54}=15.51$

3. $SD=\sqrt{\dfrac{\Sigma FD^2}{N}}=\sqrt{\dfrac{18353.88}{54}}=18.44$

2. SHORT METHOD

(1) Scores		(2) Midpoint	(3) F		(4) D	(5) FD	(6) FD^2
200–204		202.5	1		7	7	49
195–199		197.5	4		6	24	144
190–194		192.5	2		5	10	50
185–189		187.5	10	$F_g=31$	4	40	160
180–184		182.5	3		3	9	27
175–179	Average =	177.5	8		2	16	32
170–174	171.57	172.5	3		1	3 (+109)	3
165–169		167.5 (GA)	3				
160–164		162.5	4		−1	− 4	4
155–159		157.5	6		−2	−12	24
150–154		152.5	4		−3	−12	36
145–149		147.5	1	$F_l=23$	−4	− 4	16
140–144		142.5	1		−5	− 5	25
135–139		137.5	2		−6	−12	72
130–134		132.5	0		−7		
125–129		127.5	2		−8	−16 (−65)	128
			$N=54$			174	770

1. $GA=167.50$

$c=\frac{44}{54}=.8148 \quad c^2=.6639$

$C=.8148\times 5=4.07$

Average $=167.5+4.07=171.57$

2. $AD=\dfrac{\Sigma FD+c(F_l-F_g)}{N}\times\text{step}$

$=\dfrac{174+.8148(23-31)}{54}\times 5$

$=15.51$

3. $SD=\sqrt{\dfrac{\Sigma FD^2}{N}-c^2}=\sqrt{\dfrac{770}{54}-.6639}=3.687\times 5=18.44$

a correction to this guessed average (*GA*) in order to obtain the actual average. There is no set rule for guessing an average. The best plan is to take the midpoint of a step somewhere near the center of the distribution, and if possible the midpoint of that step-interval which contains the greatest frequency. In our problem the greatest *F* is on step 185–189. However, the *GA* is taken at 167.5 instead of 187.5 since the former is closer to the center of the distribution. With the question of the *GA* settled, the correction which must be applied to it to get the average is determined as outlined in the following steps:

(1) First, we fill in the *D* column, column 4. Here are entered the deviations of the midpoints of the steps measured from the *GA in units of step-interval.* Thus 172.5, the midpoint of step 170–174, deviates from 167.5, the *GA*, by 1 step-interval; and hence, a figure 1 is placed in the *D* column opposite 172.5. In like manner, 177.5 deviates 2 steps from 167.5; and accordingly, a 2 goes in the *D* column opposite 177.5. Reading on up the column from 177.5, the succeeding *D* entries are found in the same way to be 3, 4, 5, 6, and 7. The last entry, 7, is the step deviation of 202.5 from 167.5 (the actual *point* deviation, is, of course, 35).

Returning to 167.5, we find that the *D* of this point, measured from the *GA* (from itself) is 0; and hence a 0 is placed in the *D* column opposite step 165–169. Below 167.5, all of the *D* entries are negative, as all of the midpoints are less than 167.5, the *GA*. So the *D* of 162.5 from 167.5 is -1 step-interval; and the *D* of 157.5 from 167.5 is -2 step-intervals. The other *D*'s are -3, -4, -5, -6, -7, -8.

(2) The *D* column completed, we next compute the *FD* column—column 5. The *FD* entries are found in exactly the same way as in the Long Method [compare (1)]; namely, each *D* in column 4 is multiplied, or "weighted," by the appropriate *F* in column 3. Note that in the Short Method we multiply each *F* by its deviation from the *GA* in units of step-interval instead of by its actual deviation from the

average of the distribution, and that for this reason the computation of the *FD*'s is much simpler here than in the Long Method. All of the *FD*'s above (greater than) the *GA* will be positive, and all below (smaller than) the *GA* negative, since the signs of the *FD*'s depend on the signs of the *D*'s.

(3) From the *FD* column the correction is obtained as follows: The sum of the plus *FD*'s is 109; of the negative *FD*'s, −65. This makes 44 more plus *FD*'s than minus (the algebraic sum is +44) and 44 divided by 54 (*N*) equals .8148, which is the correction, "*c*," *in units of step-interval.* If we multiply *c* (.8148) by 5, the length of the step, the result is *C* (4.07), the *score correction*, or the *correction in score units.* When +4.07 is added to 167.5, the *GA*, the result is 171.57, the average. (Compare this result with the average found by the Long Method.)

A summary of the steps in the calculation of the average by the Short Method may be outlined as follows (see Table IV, 2):

(1) Organize the scores or measures into a frequency distribution.

(2) Guess an average somewhere near the center of the distribution, and preferably on the step containing the greatest frequency.

(3) Find the deviation of the midpoint of each step-interval from the *GA* in units of step-interval.

(4) Multiply or weight each step-deviation (*D*) by its appropriate *F*, i.e., by the *F* opposite it.

(5) Find the algebraic *sum* of the plus and minus *FD*'s, and divide this sum by *N*, the number of cases. This gives *c*, the *correction in units of step-interval.*

(6) Multiply *c* by the length of the step-interval to get *C*, the *score correction.*

(7) Add *C* algebraically to the guessed average to get the actual average. Sometimes *C* will be positive and sometimes negative, depending upon where the average has been guessed. The method applies equally well in either case.

If it seems to the reader that the Short Method belies its name, let him compare the calculations in columns 4 and 5 (*SM*) with the calculation of column 4 (*LM*). In spite of the extra column, the *SM* has a decided advantage over the *LM*, for as all deviations from the *GA* are in units of step-interval (whole numbers) the arithmetic is considerably easier in the latter method. In distributions containing large numbers, the calculation of the average by the *LM* becomes very laborious; and it is with such distributions that the *SM* justifies itself as a time and labor saver, rather than with distributions containing small numbers.

2. The Calculation of the *AD* by the Short Method

(A) The Calculation of the *AD* from the Average

The chief advantage in finding the *AD* by the Short Method instead of the Long Method lies in the fact (already noted in calculating the average) that in the Short Method deviations are taken from a *GA* in units of step-interval. This procedure eliminates fractions and cuts down multiplication; but at the same time it necessitates the application of a correction to the ΣFD and as a result complicates the *AD* formula. The formula for the *AD* by the Short Method is:[1]

$$AD = \frac{\Sigma FD + c(Fl - Fg)}{N} \times \text{length of step-interval.} \quad . \quad (8)$$

The term Fl in the formula refers to the sum of the F's on those steps whose midpoints are *less* (the subscript "l" means less) than the average of the distribution. The term Fg refers to the sum of the F's on those steps whose midpoints are greater (the subscript "g" means greater) than the average. In Table IV, for example, all of the midpoints from 167.5 down to 127.5, inclusive, are *less than* 171.57, the average and hence the Fl is 23. All of the midpoints from 172.5 up to 202.5, inclusive, are *greater than* 171.57; and hence the Fg is 31. It is important to remember that the Fl and the Fg

[1] This formula applies equally well to the *AD* calculated from average, median, or mode.

are always calculated from the *actual average* of the distribution (never from the guessed average) as the reference point. In consequence the 3 scores on step 165–169 whose midpoint, 167.5, is less than 171.57 are included in the Fl. A simple check on the size of the Fl and Fg is to make sure that $Fl+Fg=N$. (Note that in the present problem $23+31=54$.)

The other terms in the formula require little explanation. The c is the correction in units of step-interval. It has already been found in calculating the average (page 31) and equals .8148. The ΣFD is the arithmetic sum of the FD column, and equals 174.

If now we substitute for ΣFD, c, Fl, and Fg in formula (8), the numerator is $174+.8148(23-31)$ or 167.482. Dividing this result by $54(N)$ we obtain 3.102, the AD expressed in units of *step-interval;* and this value multiplied by 5 (the step) gives 15.51, the AD of the distribution. (Compare with the AD found by the Long Method.) Notice that it is always necessary to multiply the result given in the formula by the step-interval, since ΣFD and c are both in units of step.

Formula (8) is a relatively quick way of finding the AD of a frequency distribution. The value of the formula is somewhat limited, however, since it gives correct AD's only when c, the step-correction, is less than 1.00. In Table IV, $c=.8148$—is less than 1.00—and in consequence the formula holds, as we find on comparing the AD's given by the Long and Short Methods. One method of circumventing this limitation in the AD formula, is to make use of the fact that no matter where the GA is taken, a correction can always be calculated by means of which we can obtain the actual average. If the c so found is less than 1.00, formula (8) may be applied directly; if, however, c is larger than 1.00, we must guess another average on the same step as the actual average (which is now known) and take deviations from this "new" GA. The formula will then hold. (There is another formula for the AD which avoids the difficulty mentioned: see Kelley T. L., *Statistical Method,* p. 72ff.)

A summary of the steps in the calculation of the *AD* from the average by the Short Method may be given as follows:

(1) Find c, the correction in step-units, as shown on page 31. If c is less than 1.00:

(2) Find the arithmetic sum of the *FD*'s.

(3) Calculate the Fl: the total number of scores on steps with midpoints less than the average. Next calculate the Fg: the total number of scores on steps with midpoints greater than the average.

(4) Substitute for FD, c, Fl, Fg, N, and the step length in formula (8) to find the *AD*.

TABLE V

To Illustrate the Calculation of the *AD* from the Median by the Short Method. Data from Table II(2)

(1) Scores	(2) Midpoint	(3) F		(4) D	(5) FD
136–139	138	3		5	15
132–135	134	5		4	20
128–131	130	16	$F_g=99$	3	48
124–127	126	23		2	46
120–123	122	52		1	52
116–119	118 (GM)	49		0	
112–115	114	27	$F_l=101$	−1	−27
108–111	110	18		−2	−36
104–107	106	7		−3	−21
		$N=200$			265

$$\frac{N}{2}=100$$

$$\text{Median}=116+\frac{48}{49}\times 4=119.92$$

Guessed median $=118$ (midpoint of step 116–119)

Correction, $C=119.92-118.00=1.92$

$$c=\frac{1.92}{4}=.48$$

Applying formula: $$AD=\frac{\Sigma FD+c(F_l-F_g)}{N}\times\text{step length}$$

$$AD=\frac{265+.48(101-99)}{200}\times 4=$$

$$AD=1.33\times 4=5.32$$

(B) The Calculation of the *AD* from the Median

It is sometimes desirable to calculate the *AD* from the median instead of the average. The formula for the *AD* from the median is exactly the same as formula for *AD* from the average (see page 32). However, the scheme of the work differs in some respects from the calculation of the *AD* from the average, and hence it is illustrated in Table V for the 200 cancellation scores taken from Table II (2).

First we find the true median, 119.92, by the method outlined on pages 13–14. Next, we assume or guess a median at the midpoint of the step-interval which contains the true median, viz., at 118. Since the true median is known, the score correction, *C*, is found directly to be 1.92 by subtracting 118 from 119.92 (true median—assumed median). Then dividing 1.92 by 4, the step-interval, we obtain .48, the *correction in step-units* (*c*).

The *D*'s are taken from 118, the guessed median, and the *FD*'s are obtained (as shown in Table IV) by "weighting" each *D* by its corresponding *F*. The arithmetic sum of column 5, i.e., the ΣFD, is 265. *Fl*, the total number of scores on midpoints 118 to 106 inclusive (those less than 119.92) equals 101. And *Fg*, the total number of scores on midpoints 122 to 138 inclusive (those greater than 119.92) equals 99.

With ΣFD, *c*, *Fl*, and *Fg* known, the *AD* is now easily found by substituting these values in formula (8). The numerator becomes $265 + .48\ (101 - 99)$ or 265.96; and dividing by 200 and multiplying by 4, the step-interval, we get 5.32 as the *AD* from 119.92, the median.

3. The Calculation of the Standard Deviation (σ) by the Short Method

The calculation of the standard deviation by the Short Method is considerably less complex than the calculation of the *AD*. The formula is:

$$\sigma = \sqrt{\frac{\Sigma FD^2}{N} - c^2} \times \text{the step-interval}, \quad . \quad . \quad . \quad (9)$$

in which the ΣFD^2 is the sum of the squared deviations in units of step-intervals, taken from the guessed average, and c is the correction in units of step-interval.

An illustration of the calculation of σ by the Short Method is given in Table IV. The first step is to fill in the FD^2 column (column 6) by multiplying each D in column 4 by its corresponding FD in column 5. The process is identical with that used in the Long Method, except that the D's are all expressed in units of step-interval. This, of course, considerably simplifies the multiplication. The calculation of c has already been described on page 31. The sum of the FD^2 column (ΣFD^2) is 770, and c^2 is .6639. Applying formula (9) therefore, we get 3.687×5 or 18.44 as the σ of the distribution.

The formula for σ by the Short Method unlike the AD formula, holds good no matter what the size of the correction, c. This general applicability of formula (9) serves to increase its value.

4. The Short Method Applied to Discrete Series

We have defined a discrete series on page 2 as one in which there are real gaps. This means that in a truly discrete series each measure, instead of representing an interval on a scale as in a continuous series, is a separate and distinct value. There is, for example, a real gap between one man and two men; or between one dollar and two dollars—provided the unit of measurement in the latter case is one dollar.

Table VI illustrates the method of finding the measures of central tendency and variability for discrete measures tabulated into a frequency distribution. The data consist of the records of the number of children in 44 families of a rural community. In the first column of the table is given the number of children in the family; in the second column—under the F—the number of families of a given size. We find, for example, one family of 10 children; three of 9; four of 8, etc. Since the measures—here the children—are discrete,

TABLE VI

To Illustrate the Calculation of the Average, Median, σ, AD, and SD When Measures are Discrete

The "F" column gives the number of families containing the children listed in first column.

Measures, No. Children	F Families		D	FD	FD^2
10	1	$F_g=24$	5	5	25
9	3		4	12	48
8	4		3	12	36
7	3		2	6	12
6	5		1	5+40	5
5	8		0		
4	7	$F_l=20$	−1	−7	7
3	4		−2	−8	16
2	4		−3	−12	36
1	2		−4	−8	32
0	3		−5	−15−50	75
	$N=44$			90	292
	$\frac{N}{2}=22$				

$GA=5$

$$Q=\frac{Q_3-Q_1}{2}=\frac{6.5-3}{2}=1.75$$

$c=\frac{-10}{44}=-.23 \qquad c^2=.054$

$$AD=\frac{\Sigma FD+c(F_l-F_g)}{N}=\frac{90-.23(20-24)}{44}$$

Average $=4.77$

$AD=2.07$

Median $=5.0$

$$SD=\sqrt{\frac{\Sigma FD^2}{N}-c^2}=\sqrt{\frac{292}{44}-.054}$$

Mode $=5.0$

$SD=2.57$

$\frac{N}{2}=22$; since 22nd measure falls on 5, Median $=5$

$\frac{N}{4}=11$; since 11th measure falls on 3, $Q_1=3$.

$\frac{3N}{4}=33$; since 33rd measure falls between 6 and 7, $Q_3=6.5$.

each measure must be taken at face value, and there are, in consequence, no midpoint values for the different steps. As a result, the average being guessed at 5, D's are taken directly from this point. The FD and the FD^2 columns are calculated exactly as shown in Table IV for continuous series—the

first column is obtained by multiplying corresponding F and D values, and the second by multiplying corresponding D and FD values. Note that since the step-interval is 1, the correction c equals C directly.

If we apply the correction $-.23$ to 5, the guessed average, the average of the distribution 4.77 is obtained. This result, while mathematically correct, is obviously a rather difficult one to interpret in a practical way, however, as it is impossible for a family to have four and a fraction children. Possibly the median is a more meaningful measure. One half of the measures is 22, and counting in from the small end of the series we find that the twenty-second score falls on the frequency opposite step 5. Fractional values are, of course, really meaningless in a discrete series; and hence we must simply take 5 as being roughly the median of the distribution without any interpolation. The median family, accordingly,—and the modal family as well—may be said to contain 5 children, and on the face of it, this result seems to be of more practical value than the statement that the average number of children to a family is 4.77.

It is worth while examining further, however, exactly what is meant by the statement that the average number of children per family is 4.77. In the first place it means, of course, that the number of children in the N families examined, divided by N, gives us 4.77. But furthermore, if the families examined are actually a fair sample of all of the families in the "population" from which they are taken (see page 120), it means that if we had taken *all* of these families—or another fair sample of them—the average size of the family would have been (approximately) the same. The average, then, is a constant factor for the given population, such that, knowing the number of families in any fair sample of the population, we can multiply this number by the constant factor and obtain (approximately) the number of children in *all* of these families. Good use may thus be made of the average, therefore, even when the measures are necessarily discrete:

exactly the same kind of use that can be made of the average in the case of continuous measures.

The median, on the other hand, together with the quartiles, really breaks down in the case of discrete measures. In the example above of the families, there is actually no value which fulfills the definition of the median as such a point or value that one half of the measures exceed it, and one half fall below it. There are just 44 families in all; the median, then, would be such a point that 22 families exceeded it and 22 fell below it. Now there are 20 families falling below 5; 8 families at 5; and 16 families above 5. If we place the median exactly at 5, only 20 families instead of the required 22 fall below. And if we place the median even the least fraction above 5, the number falling below is increased by all of the families having 5 children, so that there are then 22+8 families falling below the median, or more than half. There is, in short, no median value for this series under the definition of the median which we have been using.

Sometimes, however, another definition of the median is given, namely, that it is the score or measure made by the middle individual when the individuals have been arranged in order—for scores—from least to greatest.[1] Strictly speaking, this definition also breaks down in the case of discrete measures, since there is really no sense in speaking of two or more individuals who have the same score as being *arranged* in order of magnitude, when measures are discrete. Thus the 8 families, of 5 children each, are all exactly equal as regards number of children. Of course, we might admit that in a sense, some one (any one) of these 8 families is the middle of the whole series, and since it is a family of 5 children, the median—so defined—is just 5, no more nor less. This is the median as we have used it. At best, however, it is a rough and unreliable measure.

In computing the measures of variability in a discrete series, the Q is the only one which offers difficulties. In the

[1] See discussion of midscore, page 12.

present illustration, one fourth of the measures $\left(\frac{N}{4}\right)$ is 11, and counting in from the small end of the series 11 scores, we put Q_1 on step 3 (as in the case of the median, no interpolation is made). If we check this value of Q_1 by counting in 33 scores from the large end of the distribution, we again obtain 3 as the value of Q_1. Three fourths of the measures $\left(\frac{3N}{4}\right)$ is 33; and counting in 33 scores from the small end of the series, we find that we complete—or count through—the frequency on step 6. If 11 scores are counted off from the other direction, we complete—or count through—the frequency on step 7. This puts Q_3 at either 6 or 7, and the best way out of the difficulty is to take Q_3 as roughly equal to 6.5, i.e., midway between 6 and 7. This is of course a makeshift, though even at that probably as accurate as the median or quartiles ever are in discrete series. Taking Q_1 equal to 3, and Q_3 equal to 6.5, Q is $\frac{6.5-3}{2}$ or 1.75.

The AD and σ in a discrete series are found from formulas (8) and (9) in exactly the same way as in a continuous series. For example, Fl—the number of families less than 4.77—is 22; and Fg—the number of families greater than 4.77—is 24. The AD is, therefore, $\frac{90+[-.23](20-24)}{44}\times 1$ (the step-interval) or 2.07. The σ is $\sqrt{\frac{292}{44}-.054}\times 1$ (the step-interval) or 2.57.

V. The Comparison of Groups

1. The Measurement of Relative Variability. The Coefficient of Variation

Thus far we have been dealing entirely with measures of absolute variability within the distribution, the Q, the AD, and the SD. It is sometimes desirable, however, to measure relative variability as for instance to compare the variability

of one group on two different tests, or of two or more groups on the same test. The measures of absolute variability are not sufficient in such cases as these unless the averages of the two distributions are equal or approximately equal. A problem will serve to make this clear.

A group of 50 boys works for 6 minutes on an arithmetic test and makes an average score of 20.5 with a σ of 5.24. The same group works for 10 minutes on the same test and makes an average score of 34.8 with a σ of 9.62. If we compare the σ's of these two distributions we should probably be inclined to say that the group was considerably more variable in the 10 minute period than in the 6 minute period. Despite the fact that the σ in the second period is nearly twice as large as the σ in the first period, however, this does not mean necessarily that the variability of the group has doubled with the increased time allowance (or even increased at all) for the average score has also increased from 20.5 to 34.8. In other words, the two σ's are not directly comparable as they have been measured around different central tendencies. In order to compare the relative variability of this group in the two periods it is evident, therefore, that we must have a measure which takes account *both* of the central tendency *and* the variability. Such a measure is Pearson's Coefficient of Variation, given by the formula,

$$V=\frac{100\sigma}{\text{Average}}. \qquad \qquad (10)$$

Applying this formula to the present problem we find that

$$\text{For the 6 minute period:} \quad V=\frac{5.24\times100}{20.5}=25.56.$$

$$\text{For the 10 minute period:} \quad V=\frac{9.62\times100}{34.8}=27.64.$$

Instead of being 50% as variable in the 6 minute period as in the 10, therefore, the group is seen to be actually $\frac{25.56}{27.64}$ or 93% as variable.

The coefficient of variation is especially useful in those problems in which the variability of the group under different conditions is the factor studied. As stated above, when the averages are equal the absolute variability may be compared directly.

2. The Comparison of Two Groups in Terms of Their Measures of Central Tendency and Variability

The existence of a difference between the averages or the medians of two groups does not indicate, necessarily, that there are any very marked differences in the performance of the various individuals within the two groups. An obtained difference in central tendency may mean that the person ranking lowest in the one group is better than the person ranking highest in the other; on the other hand, it may mean also that only a very small per cent of the better group is actually ahead of the poorer. For this reason in comparing groups it is not sufficient to state simply the difference between their averages or medians, for any such difference will depend for its significance largely upon the variability, or spread, within the groups compared.

Table VII will illustrate what is meant. A group of 300 boys and a group of 250 girls have been measured on the same test, and the average, median, Q and σ of each group computed. Now if we compare the central tendencies, it is clear that the average girl is 2.19 points ahead of the average boy, and that the median girl is 2.25 points ahead of the median boy. If taken alone this result might suggest a fairly definite sex difference in the given test; but before drawing this conclusion, we should compare the variability of the two groups.

A comparison of the Q's and σ's shows that the girls tend to scatter somewhat more around their central tendency than the boys. The range of scores is, however, practically the same in both groups: 100% of the boys and 92% of the girls score between 12 and 32 on the scale. Also from the quartiles it is evident that the middle 50% of the boys scored between

19 and 24 (approximately) while the middle 50% of the girls scored between 20 and 27 (approximately).

TABLE VII

COMPARISON OF TWO GROUPS IN TERMS OF CENTRAL TENDENCY, VARIABILITY, AND OVERLAPPING

Boys: Scores	F	D	FD	FD^2	Girls: Scores	F	D	FD	FD^2
28–32	15	2	30	60	32–36	20	2	40	80
24–28	68	1	68+98	68	28–32	35	1	35+75	35
20–24	128	0			24–28	73	0		
16–20	79	−1	−79	79	20–24	68	−1	−68	68
12–16	10	−2	−20−99	40	16–20	41	−2	−82	164
					12–16	13	−3	−39−189	117
	$N=300$			247		$N=250$			464
	$\frac{N}{2}=150$					$\frac{N}{2}=125$			

Boys

$GA=22.0$

$$c=\frac{-1}{300}=-.003$$

$$C=-.003\times4=-.01$$

Average $=21.99$

$$\text{Median}=20+\frac{61}{128}\times4=21.91$$

$$\left[\frac{N}{4}=75\right]Q_1=16+\frac{65}{79}\times4=19.29$$

$$\left[\frac{3N}{4}=225\right]Q_3=24+\frac{8}{68}\times4=24.47$$

$Q=2.59$

$$\sigma=\sqrt{\frac{247}{300}}\times4$$

$$=.907\times4=3.63$$

Girls

$GA=26$

$$c=\frac{-114}{250}=-.456 \qquad c^2=.208$$

$$C=-.456\times4=-1.82$$

Average $=24.18$

$$\text{Median}=24+\frac{3}{73}\times4=24.16$$

$$\left[\frac{N}{4}=62.5\right]Q_1=20+\frac{8.5}{68}\times4=20.50$$

$$\left[\frac{3N}{4}=187.5\right]Q_3=24+\frac{65.5}{73}\times4=27.59$$

$Q=3.55$

$$\sigma=\sqrt{\frac{464}{250}-.208}\times4$$

$$=1.28\times4=5.12$$

What per cent of the boys reach or exceed 24.16, the median of the girls? 217 boys score *below* 24. Step 24–28 contains 68 scores; hence there are 68/4 or 17 scores *per scale unit* on this step. $17\times.16=2.72$. $217+2.72$ or 219.72 of the boys' scores fall below 24.16, the girls' median. $300-219.72=80.28$. Accordingly, $\frac{80.28}{300}$ or 26.76%—approximately 27%—of the boys reach or exceed the median score of the girls.

Again, we find from comparing the σ's that the middle 2/3 of the boys scored between 21.99 ± 3.63, i.e., between 18 and 25 (approximately) and that the middle 2/3 of the girls scored between 24.18 ± 5.12, i.e., between 19 and 29 (approximately) on the scale. In spite of the difference in averages and medians, therefore, it is evident from the measures of variability that the boys and girls scored over almost exactly the same part of the scale.

To compare the variability of the boys as a group with that of the girls, we must compute the coefficients of variation. These are

$$\text{For Boys: } V = \frac{3.63 \times 100}{21.99} = 16.5.$$

$$\text{For Girls: } V = \frac{5.12 \times 100}{24.18} = 21.2.$$

Expressed as a per cent, the boys are $\frac{16.5}{21.2}$ or 78% as variable as the girls.

3. The Comparison of Two Groups in Terms of Overlapping

A second way of showing how alike, or unlike, two groups are in their performance on a given test is to state the amount of overlapping in the distributions of scores made by the two groups. This information serves as a valuable supplement to that secured from a comparison of central tendencies and variabilities. Overlapping is usually measured by the per cent of the one group which reaches or exceeds the median of the other. In the present problem we may compute the per cent of boys who reach or exceed the median score of the girls.

The calculation of this measure of overlapping is as follows. First, we add up the boys' scores from the small end of the distribution to find how many fall *below* 24.16, the girls' median. Two hundred and seventeen boys, $10+79+128$, score below 24, the lower limit of the step 24–28. To find how many score below 24.16, we divide the 68 scores on this

step-interval by 4 (the length of step) and multiply the result (17) by .16 in order to find how far beyond 24 we must go to reach the point 24.16. The result of this last calculation is 2.72, and accordingly a total of 217+2.72 or 219.72 of the boys' scores out of the total 300 fall *below* 24.16, the girls' median score. If we subtract 219.72 from 300, it follows that 80.28 of the boys' scores lie *above* 24.16. It is clear, then, that $\frac{80.28}{300}$ or 27% of the boys score at or beyond the girls' median. (See Table VII.)

Summarizing the results from Table VII and the discussion of the preceding paragraphs, we find that the difference between the average boy and average girl is 2.19 points in favor of the girls, and that the difference between the median boy and median girl is 2.25 points in favor of the girls. Twenty-seven per cent of the boys reach or exceed the median score of the girls; 100% of the boys and 92% of the girls score within the same limits on the scale; the middle 2/3 of the boys score between 18 and 25, and the middle 2/3 of the girls score between 19 and 29. The obvious conclusion from these data seems to be that individual differences within either group—between boy and boy or between girl and girl—are probably of more importance (because greater) than the differences between boy and girl indicated by the averages or medians taken alone.

VI. The Calculation of the Percentiles in a Frequency Distribution

We have already found it necessary in finding the quartile deviation, Q (see page 18) to calculate Q_1, the first quartile or 25th percentile, and Q_3, the third quartile, or 75th percentile. It is often very useful to know, in addition to these points, the ten decile points in the distribution as well, viz., the 10th, the 20th, the 30th, the 40th, etc., percentile points. These values are calculated in exactly the same manner as the median and the quartiles. As the 25th percentile, for example, was

found by counting off 1/4 of the scores from the small end of the distribution, and the 50th percentile (the median) by counting off 1/2 of the scores, in exactly the same way the 10th percentile is found by counting off 1/10, and the 20th percentile by counting off 2/10 of the scores from the small end of the distribution. Percentiles are of considerable value in enabling us to compare the standing of different individuals in a number of tests, or to combine the standing of the same individual in different tests (see page 278 for a fuller discussion of this).

Table VIII gives the method of calculating the percentiles in the distribution of 54 Army Alpha scores taken from Table I. The 10th percentile, 147, is located by finding 10% of 54, and counting off 5.4 scores from the small end of the distribution. In like manner, the 20th percentile, which is 2/10 or 10.8 scores from the small end of the distribution is located at 155.67. The 20th percentile *score* is taken as 155. This is due to the fact that a score of 155 in a continuous series means "155 up to 156" and consequently 155.67 falls on score 155, just as 160.25, the 30th percentile point, falls on score 160.[1] The other percentile points, and their scores, are tabulated in Table VIII.

A word should be said with regard to the calculation of the 0 and 100th percentiles. These values are the lowest and the highest scores, respectively, in the distribution. For example, we find from the original scores in Table I that the lowest score is 126 and the highest 201. Therefore, the 0 percentile falls at 126 and the 100th at 201.

Note the column in the table marked Cum. F (cumulative frequency). The entries in this column were obtained by adding the scores (the F) serially beginning with those on step 125–129: e.g., $2+0=2$; $2+2=4$; $4+1=5$, etc. From this column we can quickly tell how far we must count into the distribution in order to reach any percentile point. For example, the 70th percentile is 37.8 scores from the beginning of the distribution;

[1] This applies also to the median and the quartiles in a distribution of scores in continuous series.

TABLE VIII

To Illustrate the Calculation of the Percentiles in a Frequency Distribution

1. DATA FROM TABLE I

Scores	F	Cum. F	Percentiles	Scores
200–204	1	54	100	201
195–199	4	53	90	194
190–194	2	49	80	188
185–189	10	47	70	185
180–184	3	37	60	179
175–179	8	34	50	175
170–174	3	26	40	167
165–169	3	23	30	160
160–164	4	20	20	155
155–159	6	16	10	147
150–154	4	10	0	126
145–149	1	6		
140–144	1	5		
135–139	2	4		
130–134	0	2		
125–129	2	2		
	$N=54$			

CALCULATIONS:

10% of 54 = 5.4 $\qquad 145+\frac{.4}{1}\times 5=147$

20% of 54 = 10.8 $\qquad 155+\frac{.8}{6}\times 5=155.67$ (155)

30% of 54 = 16.2 $\qquad 160+\frac{.2}{4}\times 5=160.25$ (160)

40% of 54 = 21.6 $\qquad 165+\frac{1.6}{3}\times 5=167.67$ (167)

50% of 54 = 27 $\qquad 175+\frac{1}{8}\times 5=175.626$ (175)

60% of 54 = 32.4 $\qquad 175+\frac{6.4}{8}\times 5=179$

70% of 54 = 37.8 $\qquad 185+\frac{.8}{10}\times 5=185.40$ (185)

80% of 54 = 43.2 $\qquad 185+\frac{6.2}{10}\times 5=188.1$ (188)

90% of 54 = 48.6 $\qquad 190+\frac{1.6}{2}\times 5=194$

TABLE VIII—Continued

2. DATA FROM "A SCALE OF PERFORMANCE TESTS," BY PINTNER AND PATTERSON, PAGE 133. SCORES MADE BY 72 NINE-YEAR OLDS ON THE SUBSTITUTION TEST (IN SECONDS).

Scores (sec.)	*F*	Cum. *F*	Percentiles	Scores
80–89	1	1	100	80
90–99	2	3	90	108
100–109	5	8	80	121
110–119	5	13	70	126
120–129	13	26	60	133
130–139	9	35	50	141
140–149	6	41	40	152
150–159	11	52	30	158
160–169	5	57	20	172
170–179	3	60	10	192
180–189	4	64	0	219
190–199	3	67		
200–209	2	69		
210–219	3	72		
	$N=72$			

CALCULATIONS:

10% of 72 (90th percentile) = 7.2 $\quad 100+\frac{4.2}{5}\times 10=108.4$ (108)

20% of 72 (80th percentile) = 14.4 $\quad 120+\frac{1.4}{13}\times 10=121$

30% of 72 (70th percentile) = 21.6 $\quad 120+\frac{8.6}{10}\times 10=126.6$ (126)

40% of 72 (60th percentile) = 28.8 $\quad 130+\frac{2.8}{9}\times 10=133$

50% of 72 (50th percentile) = 36 $\quad 140+\frac{1}{6}\times 10=141.67$ (141)

60% of 72 (40th percentile) = 43.2 $\quad 150+\frac{2.2}{11}\times 10=152$

70% of 72 (30th percentile) = 50.4 $\quad 150+\frac{9.4}{11}\times 10=158.5$ (158)

80% of 72 (20th percentile) = 57.6 $\quad 170+\frac{.6}{3}\times 10=172$

90% of 72 (10th percentile) = 64.8 $\quad 190+\frac{.8}{3}\times 10=192.67$ (192)

hence it is clear from the Cum. F's that 37 scores will take us to 185—upper limit of step 180–184—and that the 70th percentile lies on step 185–189.

When once the percentile table has been drawn up, it is a relatively simple matter to find the percentile corresponding to any given score. In our problem, for instance, the man who makes a score of 177 falls on the 55th percentile—midway between the 50th (175) and the 60th (179) percentiles; while the man who scores 158 has a percentile score of 26, six tenths of the interval between the 20th percentile (155) and the 30th percentile (160). Other interpolations may be easily made in like manner.

In Table VIII (2) the percentiles have been calculated for the distribution of scores (in seconds) made by seventy-two 9-year olds on the Woodworth-Wells Substitution test.[1] As the scores are in time-units, the lowest score is the best (the quickest) performance, while the highest score is the worse (the slowest) performance. Consequently, the percentile scale is reversed: we count from the 100th percentile *down* instead of from the 0 percentile *up*. To find the 90th percentile for example, we count in 7.2 (10% of N) from 80–89 until we reach 108.4 (score 108). Counting in two tenths of N from 80–89, we reach 121, the 80th percentile. The 100th percentile is taken at 80, theoretically the fastest record; the 0 percentile at 219, the poorest record.

From the percentile table we may say that a 9-year old who completes the Substitution Test in 141 secs. has a percentile score of 50—stands at the median of the group; while a child of 9 who takes 181 secs. to complete the test stands 15th in the group—midway between the 10th percentile (192) and the 20th percentile (172).

[1] Pintner and Patterson: *A Scale of Performance Tests*, 1921, p. 133.

VII. When to Use the Various Measures of Central Tendency and Variability

The beginner in statistics is often at a loss to know which measure of central tendency or variability to use. The following summary will serve as a guide for most of the problems which the student will ordinarily meet:

1. When to Use the Average, Median, and Mode

1. Use the Average:
   (1) When each score or measure should have equal weight in determining the central tendency.
   (2) When the highest reliability is sought.
   (3) When product-moment coefficients of correlation, or measures of reliability are to be subsequently computed.
2. Use the Median:
   (1) When a quick and easily computed measure of central tendency is necessary.
   (2) When there are extreme measures which would affect the average disproportionately.
   (3) When certain scores or measures should influence the central tendency, but all that is known about them is that they are above or below the central tendency.
3. Use the Mode:
   (1) When a quick approximate measure of concentration is desired.
   (2) When only the most often recurring score is sought.

2. When to Use the Range, Q, AD, and σ

1. Use the Range:
   (1) When the data are too scant or scrappy to justify the calculation of another measure of variability.
   (2) When a knowledge of the total spread is all that is necessary.

2. Use the Q:
   (1) For a quick, inspectional measure of variability.
   (2) When there are scattered or extreme measures.
   (3) When only the concentration around the central tendency is sought.

3. Use the AD:
   (1) When it is desired to weight all deviations according to their size.
   (2) When extreme deviations should not influence the measure of variability.

4. Use σ.
   (1) When the highest reliability is desired.
   (2) When it is desired that extreme deviations influence the measure of variability.
   (3) When coefficients of correlation or measures of reliability are later to be computed.

VIII. Summary of Formulas for Finding the Measures of Central Tendency and Variability

1. Measures of Central Tendency

1. Average:

A. Long Method:

(*a*) data ungrouped:

$$\text{Average} = \frac{\Sigma(\text{Measures})}{N}. \qquad (1)$$

(*b*) data grouped:

$$\text{Average} = \frac{\Sigma(F \times M)}{N}. \qquad (2)$$

B. Short Method:

(*a*) data grouped:

Average $= GA + C$. (Algebraic.)

$$C = \frac{\Sigma(FD) \text{ (algebraic)}}{N} \times \text{length of step.}$$

2. Median:
   Arrange the measures in order of size, and count off 1/2 of the measures beginning at the small end of the series.
3. Mode:
   For Crude Mode take most frequent score, or mid-point of step with largest frequency.

2. Measures of Variability

1. Range = (largest measure) − (smallest measure).
2. Quartile Deviation:

$$Q=\frac{Q_3-Q_1}{2}, \quad \ldots \ldots \quad (3)$$

3. Average Deviation:
   A. Long Method:
      (*a*) data ungrouped:

$$AD=\frac{\Sigma D \text{ (arithmetical)}}{N}, \quad \ldots \ldots \quad (4)$$

      (*b*) data grouped:

$$AD=\frac{\Sigma FD \text{ (arithmetical)}}{N}, \quad \ldots \ldots \quad (5)$$

   B. Short Method:
      (*a*) data grouped:

$$AD=\frac{\Sigma FD+c(Fl-Fg)}{N}\times \text{length of step}, \quad \ldots \quad (8)$$

4. Standard Deviation:
   A. Long Method:
      (*a*) data ungrouped:

$$\sigma=\sqrt{\frac{\Sigma D^2}{N}}, \quad \ldots \ldots \quad (6)$$

      (*b*) data grouped:

$$\sigma=\sqrt{\frac{\Sigma FD^2}{N}}, \quad \ldots \ldots \quad (7)$$

B. Short Method:

(*a*) data grouped:

$$\sigma=\sqrt{\frac{\Sigma FD^2}{N}-c^2}\times\text{length of step}, \quad . \quad . \quad . \quad . \qquad (9)$$

5. Coefficient of Variation:

$$V=\frac{100\sigma}{\text{Average}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (10)$$

IX. Illustrative Problems

The following problems illustrate the calculation of the average, median, mode, Q, AD, and σ for continuous and discrete series. They are given as examples of the Short Method, and should be carefully reviewed by the student.

Example I

Calculation of the Average, Median, Mode, Q, AD, and SD. Step = 7

Measures	Midpoint		F		D	FD	FD^2
145–151.99	148.5		1	$F_g=34$	6	6	36
138–144.99	141.5		1		5	5	25
131–137.99	134.5		2		4	8	32
124–130.99	127.5		2		3	6	18
117–123.99	120.5		3		2	6	12
110–116.99	113.5		10		1	10+41	10
103–109.99	106.5	Av. =	15				
96–102.99	99.5	106.26	14	$Fl=25$	−1	−14	14
89– 95.99	92.5		6		−2	−12	24
82– 88.99	85.5		3		−3	− 9	27
75– 81.99	78.5		2		−4	− 8−43	32
			$N=59$			84	230

$$\frac{N}{2}=29.5$$

$GA=106.5$

$c=-\frac{2}{59}=-.034 \qquad c^2=.001$

$C=-.034\times7=-.238$

Average $=106.5+(-.238)=106.26$

Median $=103+\frac{4.5}{15}\times7=105.10$

Mode $=106.50$

$\left[\frac{N}{4}=14.75\right]Q_1=96+\frac{3.75}{14}\times7=97.875$

$\left[\frac{3N}{4}=44.25\right]Q_3=110+\frac{4.25}{10}\times7=112.975$

$Q=7.55$

$AD=\frac{84+(-.034)[25-34]}{59}\times7$

$AD=10.00$

$\sigma=\sqrt{\frac{230}{59}-.001}\times7$

$\sigma=1.97\times7=13.79$

Example II

Calculation of Average, Median, Q and SD. Step = 1

Scores	F	D	FD	FD^2
22–22.9	1	12	12	144
21–21.9	7	11	77	847
20–20.9	16	10	160	1,600
19–19.9	35	9	305	2,745
18–18.9	81	8	648	5,184
17–17.9	172	7	1,204	8,428
16–16.9	330	6	1,980	11,880
15–15.9	600	5	3,000	15,000
14–14.9	1,031	4	4,124	16,496
13–13.9	1,793	3	5,379	16,137
12–12.9	2,572	2	5,144	10,288
11–11.9	2,951	1	2,951 +24,984	
10–10.9	3,187	0		
9– 9.9	3,319	−1	−3,319	3,319
8– 8.9	2,891	−2	−5,782	11,564
7– 7.9	2,149	−3	−6,447	19,341
6– 6.9	1,315	−4	−5,260	21,040
5– 5.9	684	−5	−3,420	17,100
4– 4.9	302	−6	−1,812	10,872
3– 3.9	112	−7	− 784	5,488
2– 2.9	38	−8	− 304	2,432
1– 1.9	10	−9	− 90 −27,218	810
	$N = 23{,}596$		−2,224	180,715

$$\frac{N}{2} = 11{,}798$$

$GA = 10.5$

$$c = \frac{-2234}{23{,}596} = -.09 \qquad c^2 = .008 \qquad \sigma = \sqrt{\frac{180{,}715}{23{,}596} - .008} \times 1$$

$C = -.09$

$\text{Average} = 10.41 \qquad \sigma = 2.77$

$$\text{Median} = 10 + \frac{978}{3187} \times 1 = 10.31$$

$$\left[\frac{N}{4} = 5{,}899\right] Q_1 = 8 + \frac{1289}{2891} \times 1 = 8.45 \qquad Q = 1.92$$

$$\left[\frac{3N}{4} = 17{,}697\right] Q_3 = 12 + \frac{739}{2572} \times 1 = 12.29$$

Example III

Calculation of Average, Median, Mode, Q, AD, SD, for Discrete Series
Step = 1

Measures		F		D	FD		FD²
21		2	F_l	−4	−8		32
22		1		−3	−3		9
23		4		−2	−8		16
24		9		−1	−9	−28	9
25		21		1	11		11
—	Average = 25.036			2	12		24
26		11	F_g	3	3		9
27		6		4	4	+30	16
28		1					
29		1			58		126

$N = 56$

$\frac{N}{2} = 28$

$GA = 25$

$c = \frac{2}{56} = .036 \qquad c^2 = .001$

Average = 25.04

Median = 25

Mode = 25

$\left[\frac{N}{4} = 14\right] \qquad Q_1 = 24$

$\left[\frac{3N}{4} = 42\right] \qquad Q_3 = 26$

$$AD = \frac{58 + .036(37-19)}{56} \times 1$$

$AD = 1.05$

$$\sigma = \sqrt{\frac{126}{56} - .001} \times 1$$

$\sigma = 1.50$

$Q = 1.0$

PROBLEMS

1. Tabulate the following scores into three frequency distributions, using class-intervals of 3, 5, and 10 units respectively.

Scores made on the Thorndike Entrance Examination by 100 applicants for admission to Columbia College. (From Sommerville, R. C.: *Physical, Motor and Sensory Traits*, Archives of Psychology, 75, 1924.) Note:—Fractions have been dropped.

63	80	75	90	81	83
78	81	83	83	89	98
46	90	103	81	71	93
82	78	86	85	73	83
74	86	84	72	63	76
103	78	85	81	105	94
78	101	76	96	74	75
86	65	80	81	98	56
103	90	92	85	78	73
87	75	102	58	78	95
73	73	73	96	83	110
95	90	87	86	96	98
82	86	70	70	95	71
89	86	85	72	94	92
73	84	79	74	88	72
92	86	93	84	50	
85	76	82	99	91	

2. The following distributions represent the scores made on a logical memory test by two racial groups, A and B.
   (1) Find the average, median, Q and SD of each distribution.
   (2) What per cent of group A reaches or exceeds the median of group B?
   (3) Compare the relative variability of the two groups by means of their coefficients of variation.

Scores	Group A	Group B
79–83	6	8
74–78	7	8
69–73	8	9
64–68	10	16
59–63	12	20
54–58	15	18
49–53	23	19
44–48	16	11
39–43	10	13
34–38	12	8
29–33	6	7
24–28	3	2
	$N=128$	$N=139$

3. Compare the 30th, 60th, and 90th percentile scores in Group A [problem (2)] with the corresponding percentile scores in Group B.

4. The following problems are given for the purpose of affording practice in finding measures of central tendency and measures of variability. In every case where the Average, *AD*, or *SD* is to be found, use the Short Method.

(1) Find the Average and *SD*.

Scores	*F*
70–71	2
68–69	2
66–67	3
64–65	4
62–63	6
60–61	7
58–59	5
56–57	4
54–55	2
52–53	3
50–51	1
	$N=39$

(2) Find the Median and *AD* (from the Median.)

Scores	*F*
90–94	2
85–89	2
80–84	4
75–79	8
70–74	6
65–69	11
60–64	9
55–59	7
50–54	5
45–49	0
40–44	2
	$N=56$

(3) Find the Average, *AD*, and *SD*.

Scores	*F*
120–122	2
117–119	2
114–116	2
111–113	4
108–110	5
105–107	9
102–104	6
99–101	3
96–98	4
93–95	2
90–92	1
	$N=40$

(4) Find the Average and *SD*. (Discrete Series.)

Scores	*F*
80	1
79	3
78	3
77	6
76	8
75	7
74	3
73	4
72	2
71	1
	$N=38$

(5) Find the Median and Q.

Scores	F
100–109	5
90–99	9
80–89	14
70–79	19
60–69	21
50–59	30
40–49	25
30–39	15
20–29	10
10–19	8
0–9	6
	$N=162$

(6) Find the Average, Median and SD.

Measures	F
80–84	8
75–79	14
70–74	19
65–69	24
60–64	29
55–59	27
50–54	26
45–49	28
40–44	20
35–39	15
30–34	10
	$N=220$

Answers

		Group A	Group B
2. (1)	Average	53.88	56.21
	Median	52.70	56.64
	Q	9.64	9.90
	SD	13.82	13.73

(2) 39% of Group A reaches or exceeds the median of Group B

(3) Coefficient of Variation, Group A = 25.64; Group B = 24.43; Group B is 95.3% as variable as Group A.

		Group A	Group B
3.	30th percentile score	46	49
	60th percentile score	56	60
	90th percentile score	74	75

4. (1)	Average = 61.26	$SD =$ 4.99	
(2)	Median = 67.27	$AD =$ 8.97	
(3)	Average = 106.5	$AD =$ 5.55	$SD = 7.23$
(4)	Average = 75.66	$SD =$ 2.11	
(5)	Median = 55.67	$Q = 16.41$	
(6)	Average = 57.0	Median = 57.04	$SD = 13.17$

CHAPTER II

GRAPHIC METHODS AND THE NORMAL CURVE

I. The Graphic Representation of the Frequency Distribution

We learned in the last chapter how scores or other measures of capacity may be organized and condensed into the tabular arrangement called a frequency distribution. In addition we found how such arrangement aids us in calculating measures of central tendency and variability, and, in general, gives us a better idea of the facts as a whole. Still further aid in analyzing numerical data may be secured by a graphic or pictorial treatment of our material. The advertiser has long recognized the power of the illustration to catch the eye and hold the attention where the most careful array of statistics fails. And in like manner, the statistician, through the medium of diagrams and graphs, attempts to utilize the attention-getting power of visual presentation and at the same time to translate numerical facts—often abstract and difficult of interpretation—into a more concrete and understandable form.

There are three methods of representing graphically—i.e., of "plotting"—measures which have been grouped into a frequency distribution. The first method gives the Frequency Polygon; the second the Histogram or Column Diagram; and the third, the Ogive, or cumulative frequency graph. These will be considered in order.

1. The Frequency Polygon

Before outlining the method of constructing a frequency polygon, it might be well to review briefly the simple algebraic principles which apply to all graphical representation of

numerical data. Graphing or plotting is done with reference to two lines or "coordinate axes," the one the vertical or Y-axis, the other the horizontal or X-axis. These basic lines are perpendicular to each other, the point where they intersect being called O, or the *origin*" (see Diagram II). To locate or "plot" a point "P" whose coordinates are $x=4$, and $y=3$, we go out from the origin 4 units on the X-axis, and up from the origin 3 units on the Y-axis, and, where the perpendiculars to these points intersect, locate the point P (see Diagram II). In like manner, any point whose x and y values are known can be located with reference to OY and OX, the coordinate axes. Distances measured along the X-axis are commonly called *abscissas*, and distances along the Y-axis *ordinates*.

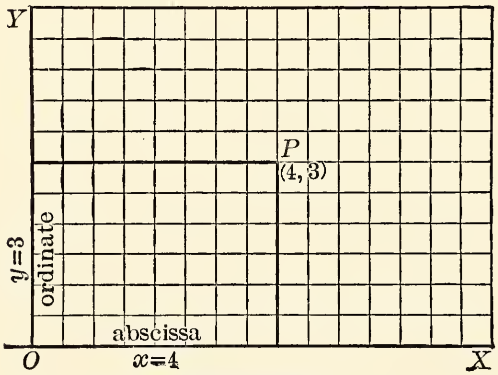


DIAGRAM II
THE USE OF COÖRDINATE AXES X AND Y.

We may now show how these principles of graphing apply to the construction of the frequency polygon shown in Diagram III (1). This graph pictures the frequency distribution of Table I. The limits of the step-intervals (the abscissas) are laid off at regular intervals along the base line (the X-axis) from the origin; and the frequencies within each interval (the ordinates) are measured off on a scale along the Y-axis. There are 2 scores on the first step, 125–129 (see Table I). To represent these on our diagram, we go out on the X-axis to 127.5—midway between 125 and 130—and up 2 Y-units. Here we locate the first point. The frequency on the next step-interval, 130–134 is 0; hence the second point falls midway between 130 and 135 directly on the X-axis. The 2 scores on step 135–139, the 1 score on step 140–144, and the frequency on each succeeding step is, in every case, represented

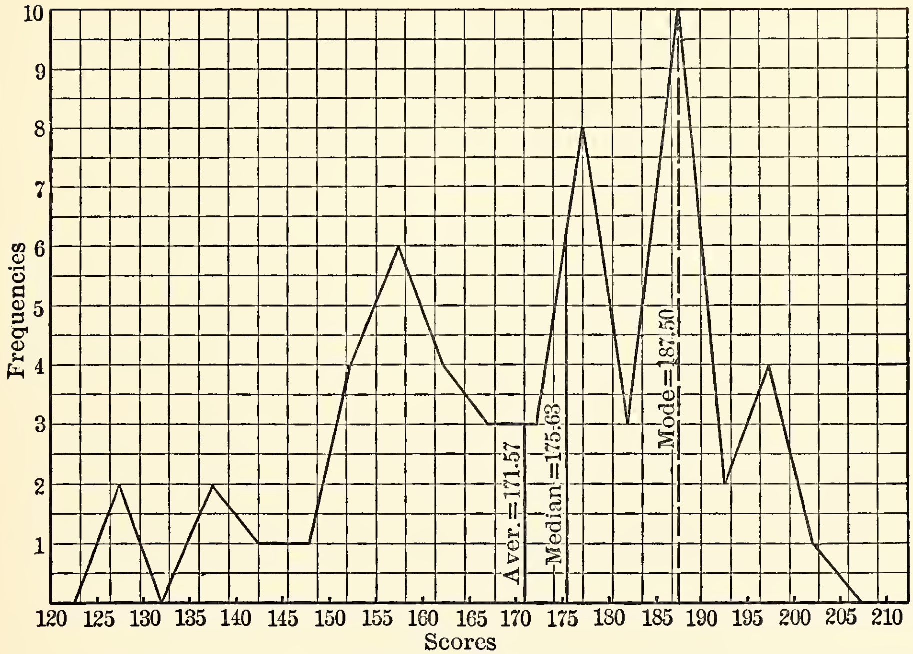


DIAGRAM III (1)

FREQUENCY POLYGON PLOTTED FROM DISTRIBUTION OF 54 SCORES IN TABLE I

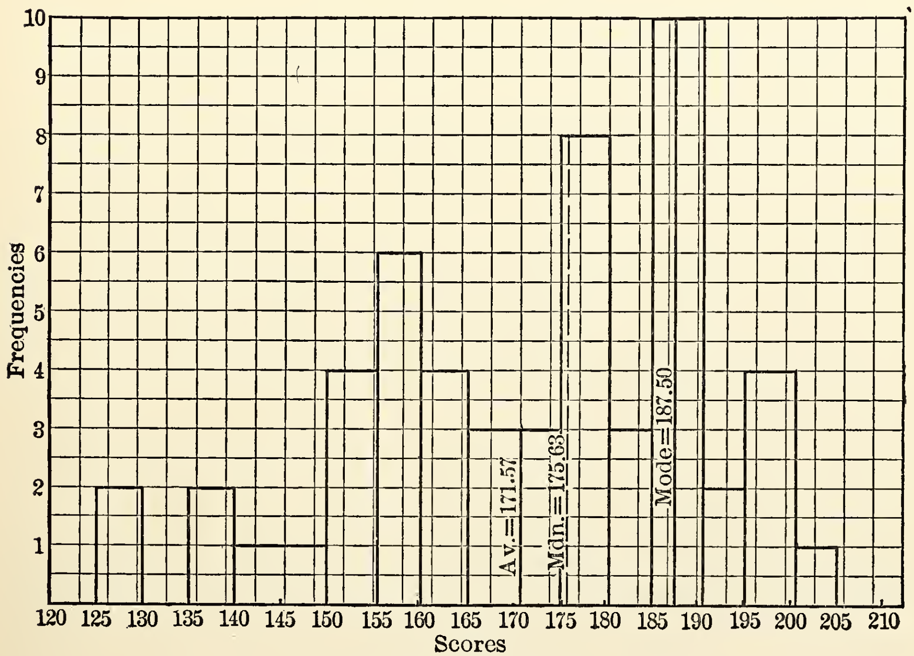


DIAGRAM III (2)

HISTOGRAM PLOTTED FROM DATA IN TABLE I.

by a point the specified number of scores (Y-units) above the X-axis, and midway between the upper and lower limits of the step on which it lies. It is important to remember in plotting a frequency polygon that the *midpoint* of the step is always taken to represent *all* of the scores within that interval. The heights of the ordinates at the different midpoints represent the frequencies within the intervals.

When all of the points have been located they are joined in regular order to give the outline of the frequency polygon shown in Diagram III (1). In order to complete the figure, note that the step next below the lowest (125–129) and the step next above the highest (200–204) are included on the X-scale. The frequency of each of these steps is taken as 0; and in consequence the frequency polygon begins and ends on the X-axis.

The distance taken to represent a step-interval on the X-axis will usually depend on the width of the cross section paper used and on the number of steps in the distribution. No general rule can be given for the choice of an X-unit: nor for the choice of the unit taken to represent 1 score on the Y-axis. The length of the diagram, and the maximum frequency on any given step (as, for example, the 10 scores on step 185–189) will generally serve to indicate within what practical limits the Y-unit must be selected. After plotting several polygons, the student will soon discover that a too-long Y-unit exaggerates the changes in the distribution from step to step, while a too-short Y-unit makes the graph too flat. In like manner, a too-long X-unit tends to stretch out the polygon, while a too-short X-unit crowds the separate points on the frequency surface and makes comparisons difficult.

The total frequency (N) of the distribution is represented by the *area* of the polygon: that is, by the area between the boundary or frequency surface and the base line. The area of any given interval cannot be taken as proportional to the number of cases within the interval, however, because of the

numerous irregularities in the distribution, and consequently of the frequency surface.

To show the position of the average, median, and mode on the graph, we must first locate these values on the X-axis, and then erect perpendiculars as shown in the diagram. Note that the mode is easily located as the highest point on the frequency surface.

The steps involved in constructing a frequency polygon may be summarized as follows:

1. Draw two straight lines perpendicular to each other, the vertical line near the left side of the paper, the horizontal line near the bottom. Call the vertical line—the Y-axis—OY, and the horizontal line—the X-axis—OX. Put the O where the two lines intersect. This point is called the origin.
2. Lay off the step-intervals of the frequency distribution at regular intervals along the X-axis. Begin with the lower limit of the step next below the lowest as the origin, and end with the upper limit of the step next above the highest. Label the successive X-points with the step limits. Select as the X unit a distance which will permit all of the steps to be represented on the one graph.
3. Mark off on the Y-axis successive unit distances to represent the scores on the different steps. Choose a scale which will permit the maximum frequency to be represented on the graph.
4. From the midpoint of each step-interval on the X-axis, go up in the Y direction a distance equal to the number of scores on the step. Place a point here.
5. Join the points plotted in (4) with straight lines to give the frequency polygon.

2. The Histogram or Column Diagram

A second method of representing a frequency distribution graphically is to construct a histogram or column diagram. This type of graph is illustrated in Diagram III (2), with the same distribution of scores represented by the frequency polygon in Diagram III (1). The two graphs are constructed

in much the same way with this important difference: that whereas, in a frequency polygon, all of the scores within a given interval are represented by the midpoint of that interval, in the histogram the assumption is made that all of the scores within an interval are spread uniformly over the *entire* interval. For this reason, the measures within any given interval in a histogram are represented by a rectangle constructed with base equal to the length of the step-interval, and altitude equal to the number of measures within the interval. Thus [see Diagram III (2)] the 2 scores on step 125–129 are represented by a rectangle with base equal to the length of step-interval on the X-axis, and altitude equal to 2 units measured off on the Y-axis. As there are no scores within the next interval 130–134, no rectangle is drawn here. The altitudes of the other rectangles vary with the number of scores on the intervals. When the same number of scores occur on two (or more) adjacent steps, as in the intervals from 140 up to 145 and from 145 up to 150, the base of the rectangle covers two (or more) intervals on the X-axis. The highest rectangle is, of course, that which has the step 185 up to 189 as its base and 10, the maximum frequency, as its altitude. In selecting scales for the X- and Y-axes, the same considerations as to numbers of intervals, size of paper, maximum frequency, etc., noted under the frequency polygon, must be observed.

Although in a histogram each step-interval is represented by a separate rectangle, it is not necessary to project the sides of these different rectangles to the base line, as shown in Diagram III (2), as the rise and fall of the boundary line showing the increase or decrease in the number of scores from step to step is usually the important fact to be brought out. As in the frequency polygon, the total frequency (N) is represented by the *area* of the histogram. In contrast to the frequency polygon, however, the area of *each rectangle* in a histogram is directly proportional to the number of measures in the interval, so that we have in the column diagram an accurate picture of the number of scores falling on each step.

In order to make easier a comparison of the two types of frequency graph, the distribution of Table III is plotted in Diagram IV, on the same coordinate axes, both as a frequency polygon and a histogram. The increased number of cases and the more symmetrical distribution of scores make both

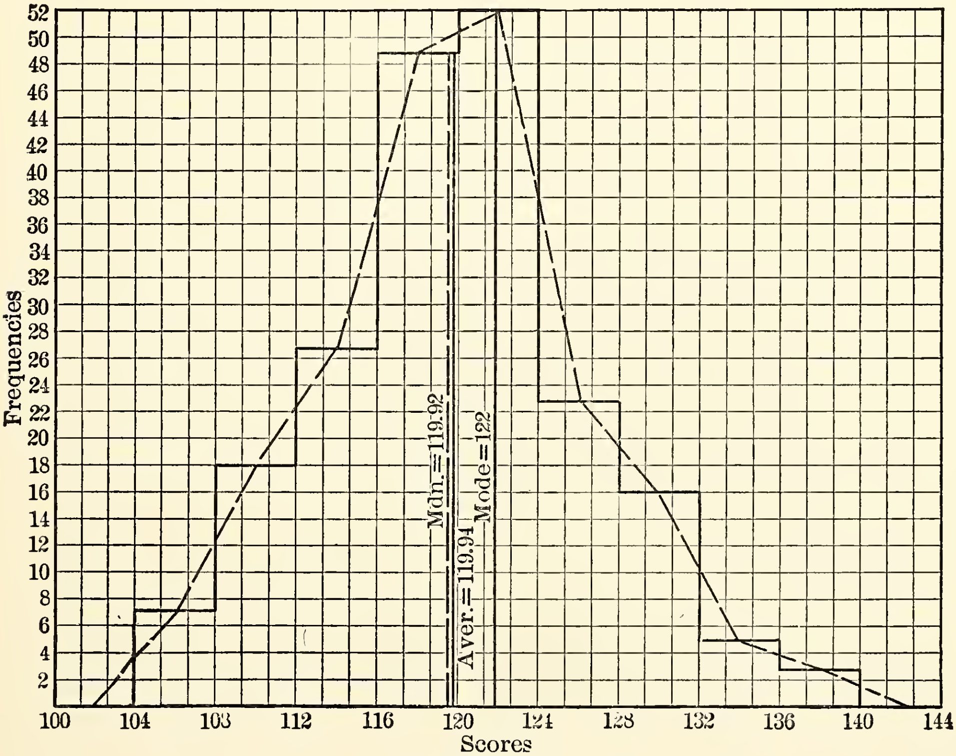


DIAGRAM IV

Plotting of Frequency Polygon and Histogram.
[Data from Table III (2)].

of these graphs more regular in appearance than the graphs of Diagram III.[1]

The question of when to use the frequency polygon and when to use the histogram cannot be answered, unfortunately, by giving a general rule which will cover all cases. The frequency polygon is less exact than the histogram in that it does not represent accurately—i.e., in terms of area—the

[1] Other examples of frequency polygons and histograms may be found on page 75.

number of measures on the successive step-intervals. For comparing two or more distributions plotted on the same diagram, however, the frequency polygon is probably the more useful, since the many vertical lines in the histogram often coincide. Both the histogram and the frequency polygon tell the same story, and both are useful in enabling us to show in a graphic fashion whether the scores of a group distribute uniformly over the scale, or whether they pile up at the low or the high end. Not only information with regard to the group but information with regard to the test may be thus secured. If a test is too easy, the scores will fall disproportionately at the high end of the scale; if too hard at the low end. If the test is neither too hard nor too easy, the scores will tend to be symmetrically distributed, a few individuals scoring high, a few low, and the majority scoring somewhere near the middle of the scale. In this last case, the frequency polygon or histogram approximates the "ideal" or normal frequency distribution (see page 76).

3. The Ogive

The ogive, or cumulative frequency graph, is a third way of representing a frequency distribution by means of a diagram. Before we can plot an ogive, the scores of the distribution must first be added serially or cumulated, as shown in Table IX for the two distributions taken from Table II (1 and 2). (These two distributions have already been used to illustrate the frequency polygon and histogram in Diagrams III and IV.) Note that the first two columns in Table IX are exactly the same as in any frequency distribution, but that in the third column the scores have been "accumulated" successively from the low end of the distribution as described on page 46. The last cumulative score is, of course, equal to N.[1]

[1] Cumulative distributions are useful also in telling quickly how many in a group scored above or below a certain point on the scale. In Table IX, for example, we read that 10 men in the group made Alpha scores below 155, 47 below 190, etc.

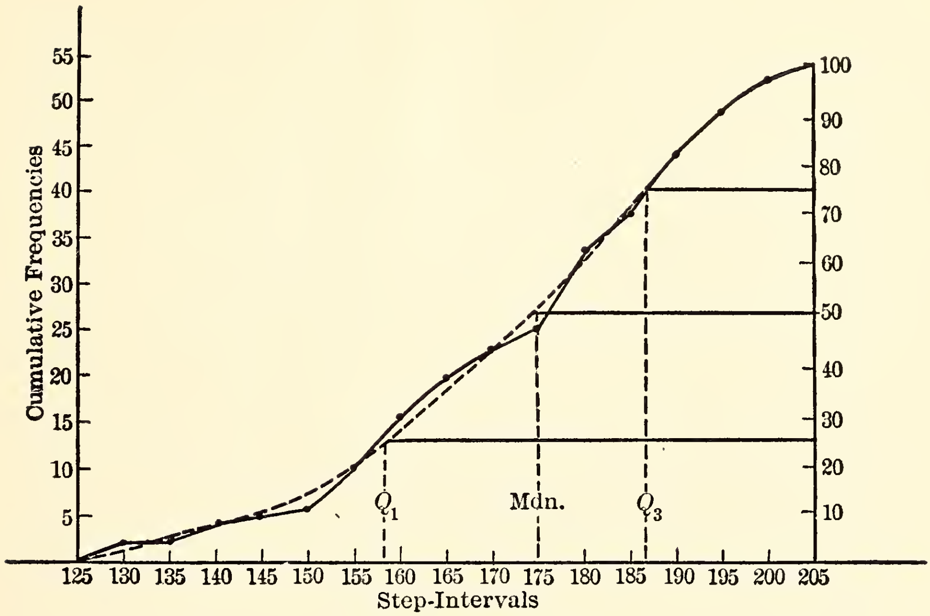


DIAGRAM V (1)

Ogive Curve. Data from Table II (1).

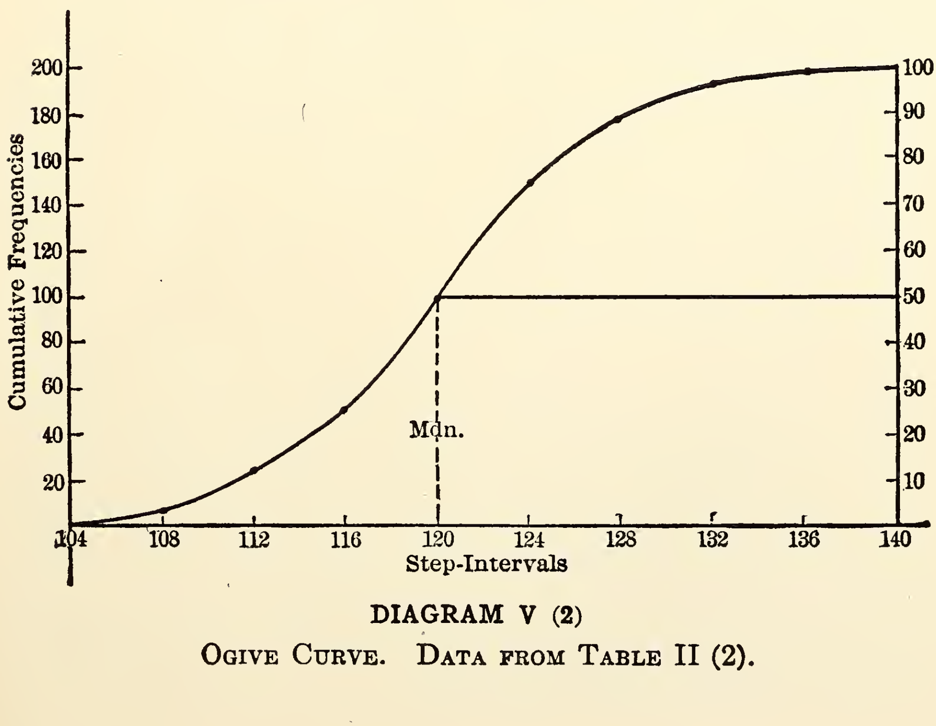


DIAGRAM V (2)

Ogive Curve. Data from Table II (2).

The two ogives which represent the distributions of Table IX are shown in Diagram V (1 and 2). Consider first the ogive of the 54 Alpha scores shown in (1). The step-intervals of the distribution have been laid off along the X-axis, and successive distances equal to the total number of scores in the distribution (here 54) have been laid off on the Y-axis. It will be remembered in plotting the frequency polygon that the frequency of each step was taken at the *midpoint* of the step-interval; in constructing an ogive, however, each cumulative

TABLE IX

CUMULATIVE FREQUENCIES OF THE TWO DISTRIBUTIONS IN TABLE II

(For Plotting the Ogives of Diagram V)

(1) Measures	F	Cum. F	(2) Measures	F	Cum. F
200–204	1	54	136–139	3	200
195-199	4	53	132–135	5	197
190–194	2	49	128–131	16	192
185–189	10	47	124–127	23	176
180–184	3	37	120–123	52	153
175–179	8	34	116–119	49	101
170–174	3	26	112–115	27	52
165–169	3	23	108–111	18	25
160–164	4	20	104–107	7	7
155–159	6	16			
150–154	4	10		$N=200$	
145–149	1	6			
140–144	1	5			
135–139	2	4			
130–134	0	2			
125–129	2	2			
	$N=54$				

frequency must be plotted at the *upper limit* of the step on which it falls. The first point on the curve, for example, is 2 Y-units (the cumulative frequency on step 125–129) above 130; the second point is 2 Y-units above 135, the third, 4 Y-units above 140, and so on to the last point which is 54 Y-units above 205. The plotted points are joined in order to give the ogive. Note that the curve begins at 125 on the X-axis, and ends at 205 just 54 Y-units *above* the X-axis.

Because the sample is small and the distribution of scores unsymmetrical, the ogive in (1) is somewhat jagged in outline. To eliminate such irregularities as these and to facilitate later computations, we often "smooth" an ogive by sketching in a smooth curve through as many of its points as possible. The dotted line in Diagram V (1) shows the result of this smoothing process. If the sample is large, and the measures well distributed, smoothing is often unnecessary [see Diagram V (2)].

The ogive in Diagram V (2) has been plotted from the distribution in Table IX (2), as described above. It offers no new difficulties and need not be considered in any detail. Note that the curve begins at 104, the lower limit of the first step, and ends at 140, the upper limit of the last step on the scale; also that the cumulative F's, 7, 25, 52, etc., have all been plotted at the upper limits of their respective step-intervals. This ogive does not require any smoothing as the distribution which it represents is very symmetrical.

The ogive has been less frequently used by workers in experimental psychology and education than either the frequency polygon or the histogram, and is probably somewhat more difficult for the general reader to interpret. It has, however, several distinct advantages. In the first place, unlike the other frequency graphs, the shape of the ogive remains practically the same when the size of the step-interval varies. Furthermore, while the frequency polygon and histogram cannot be compared unless the step-intervals are the same, this restriction does not apply to the ogive.

Probably the chief value of the ogive to the student of mental measurement lies in the relative ease with which percentile values may be calculated from the curve. The method of getting these values is illustrated in Diagram V (1 and 2). First, a perpendicular is erected on the X-axis at the upper limit of the last step-interval, and continued until it reaches the curve. (In the first ogive this perpendicular will be erected at 205.) Next, this line between the curve and the

X-axis is divided into 10 equal parts (by means of a compass or mm. rule) and the points of division labeled 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 (the 100 point lies on the curve, the 0 point on the X-axis). These points are used to locate the 10 decile points in the distribution. To find the second decile, or 20th percentile, for example, we draw a line from the second point, i.e., from 20, parallel to the X-axis, and where this line cuts the curve, drop a perpendicular to the X-axis.

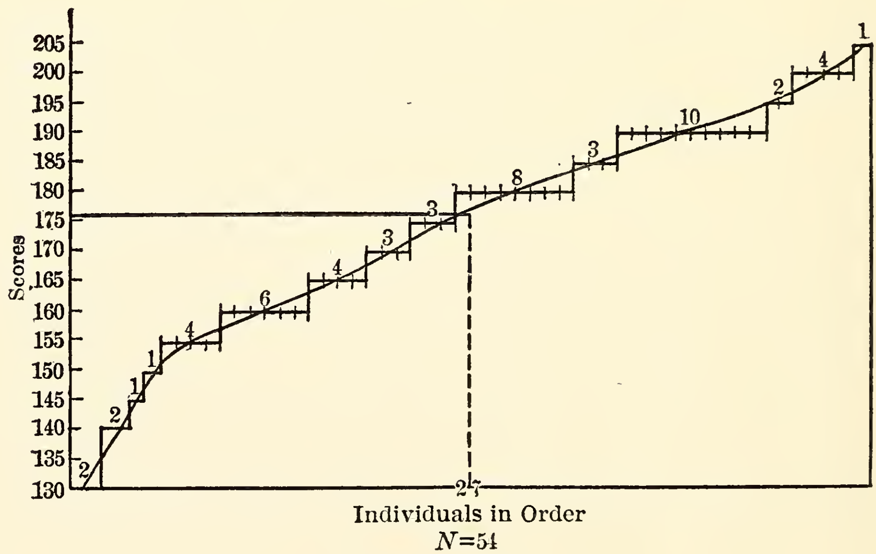


DIAGRAM VI

Another Way of Constructing an Ogive. The Individuals are Arranged in Order Along the Baseline, Each Man's Score Being Marked Off on the Ordinate Above Him.

This perpendicular locates the 20th percentile on the X-scale. The other percentiles and quartiles may be found in the same way. Notice in ogive (1) that the 0 percentile is 125—theoretically the lowest score in the distribution—and that the 100th percentile is 205—theoretically the highest score in the distribution.

The student should compare the percentile values obtained from the ogive with the same values as calculated in Table VIII (1). Due to the greater smoothness of the curve, the

percentiles obtained from ogive (2) will be more accurate than those got from the ogive (1).

The accuracy with which we are able to obtain the percentiles graphically will depend, in general, on the accuracy with which the points of the curve have been plotted, the fineness of the scale, the number of cases, and the symmetry of the distribution.

Another way of constructing an ogive is shown in Diagram VI, with the data of Table IX (1). Imagine the 54 individuals in the distribution arranged along the baseline according to the size of their scores, the score of each man being marked off on the ordinate above him. When these points are joined by straight lines, we have a series of rectangles of the histogram type, the base of each rectangle representing the number of men making the given score, the height of each rectangle representing the size of the score. A smooth curve may be sketched through (or as near as possible to) the midpoint of the upper base of each rectangle—as shown in the diagram—to give an ogive curve. From this ogive, percentiles may be easily found. To get the median, for example, we erect a perpendicular at 27 $\left(\frac{N}{2}\right)$ on the X-axis, and draw a line through the point where this perpendicular cuts the curve parallel to the X-axis to locate the median approximately at 175 on the Y-scale. The quartiles and the percentile points may be found in exactly the same manner.

II. Other Uses of Graphical Methods—the Comparative Line Graph

Many problems in mental measurement, especially those which involve the measurement of changes attributable to growth, learning, practice, etc., readily lend themselves to graphical treatment. Diagram VII illustrates several such problems, in which the data are represented by "line graphs." As in all graphs hitherto considered, the measures are plotted

with reference to the coordinate axes, OY and OX, the coordinates of a plotted point being its abscissa or X-distance, and its ordinate, or Y-distance.

Figure 1 illustrates the "age" or "growth" curve. It

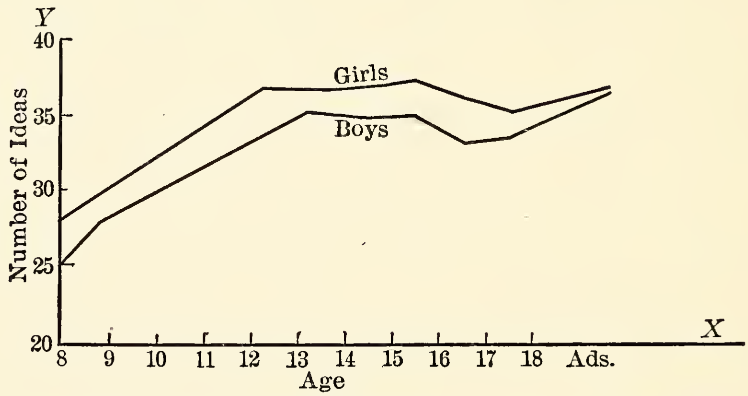


Fig. 1.—Logical memory. Age is represented on X-line (horizontal); score, e.g., number of ideas remembered, on Y-line (vertical). (After Pyle.)

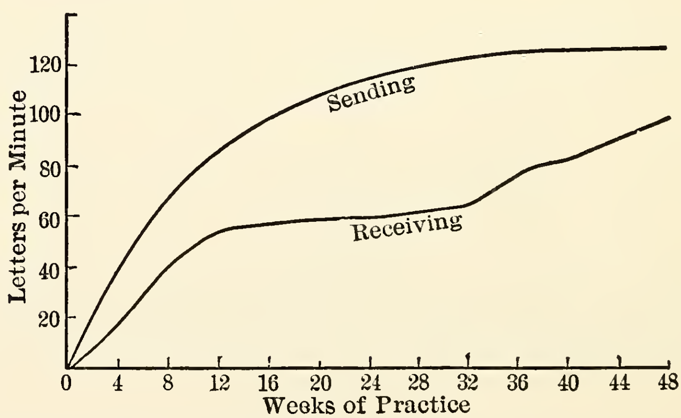


Fig. 2.—Improvement in telegraphy. Weeks of practice on X-lines; number of letters per minute on Y-line. (After Bryan and Harter.)

DIAGRAM VII

Comparative Line Graphs.

represents the growth in logical memory (for a connected passage) in boys and girls from 8 to 18 years old.

Figure 2 illustrates the "learning" or "practice" curve. It shows the improvement in sending and receiving telegraphic messages, resulting from successive trials at the same task

over a period of weeks. Improvement is measured in terms of the number of letters sent or received per minute.

Figure 3 is a "performance" or "practice" curve. It represents 25 successive trials with the hand dynamometer

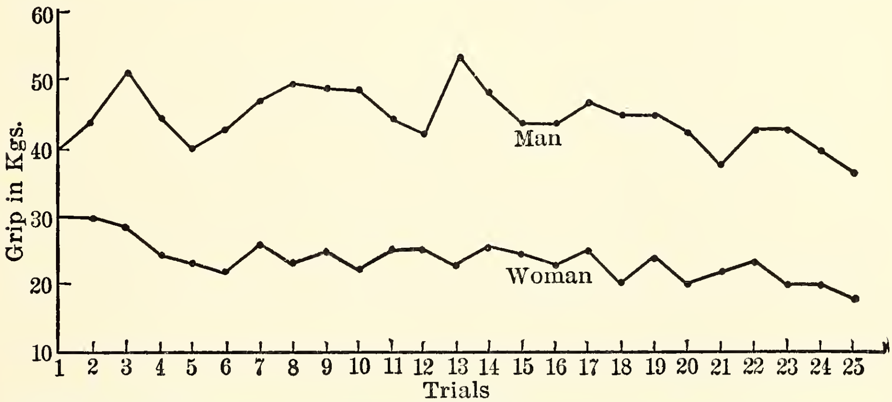


FIG. 3.—Hand dynamometer readings in kilograms for 25 successive grips at intervals of 10 seconds. Two subjects, a man and a woman.

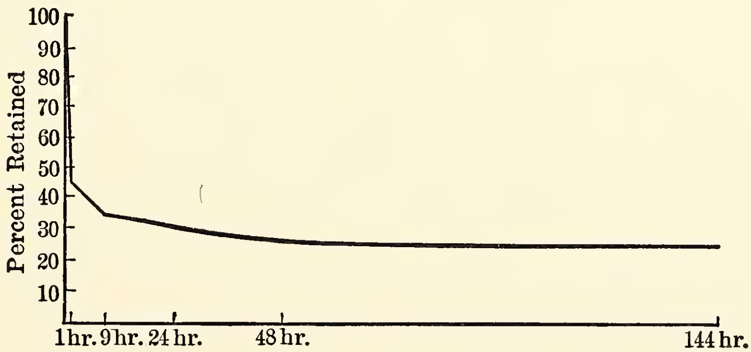


FIG. 4.—Curve of forgetting. The numbers on base line give hours elapsed from time of learning; numbers along Y-axis give per cent retained. (After Ebbinghaus.)

DIAGRAM VII

COMPARATIVE LINE GRAPHS.

by one man and one woman. Note that the successive trials are laid off on the X-axis, and the strength of grip (in kgs.) on the Y-axis. Graphs like these are useful in enabling us to compare individuals or groups at various stages in the test or performance. They also enable us to study the effect of fatigue with successive trials.

Figure 4 shows the well-known "curve of forgetting" (or

retention). It represents memory retention, as measured by the percentage of the original material retained after the passage of different time intervals. The time intervals between relearning are laid off on the X-axis; the per cent retained, as shown by the relearning, on the X-axis.

III. The Normal Probability Curve

In Diagram VIII are shown four graphs—two frequency polygons and two histograms—which represent frequency distributions of data drawn from anthropometry, psychology, and meteorology. It is at once apparent that all of these graphs have the same general form—the measures are concentrated closely around the center, and taper off from the central high point, or crest, equally to right and left. In general we find relatively few measures at the "low" score end of the scale; an increasing number up to a maximum at the midposition, and a progressive falling off as we go toward the "high" score end of the scale. If we divide the area *under* each curve (the area between the curve and the X-axis) by a line drawn perpendicularly through the central high point to the base line, the two parts will be practically similar in form and equal in area. This results from the fact that each curve shows almost perfect bilateral symmetry. The perfectly symmetrical curve, or frequency surface, to which all of the figures in Diagram VIII approximate, is shown in Diagram IX. This bell-shaped curve is called the Normal Probability Curve, or simply the Normal Curve, and is of the greatest value in psychological measurement. An understanding of its characteristics is essential to the student of experimental psychology and measurement; and consequently the rest of this chapter will be concerned with the study of the properties and uses of the Normal Curve.

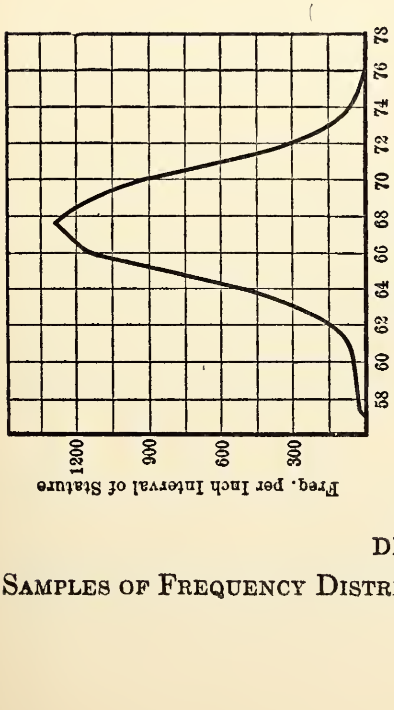


FIG. 1.—Statures of 8585 adult males born in British Isles. (After Yule, page 89.)

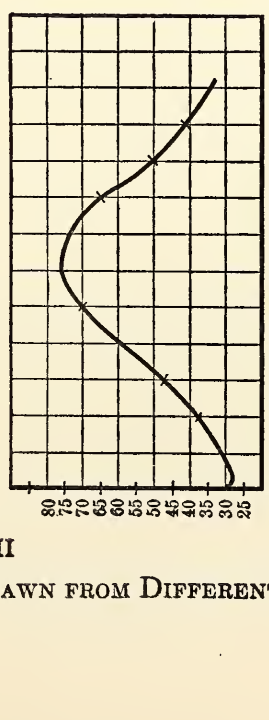


FIG. 2.—Mean monthly temperatures, New York City, Jan.-Dec., average for 47 years. (After Kelley, page 28.)

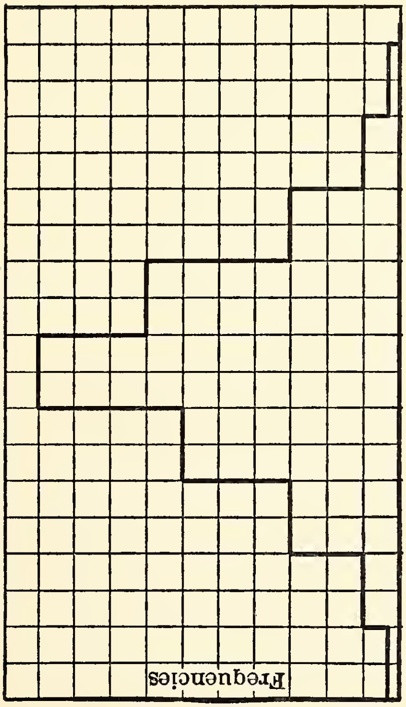


FIG. 3.—*IQ*'s of 905 unselected children, 5–14 years old. (After Terman, page 66)

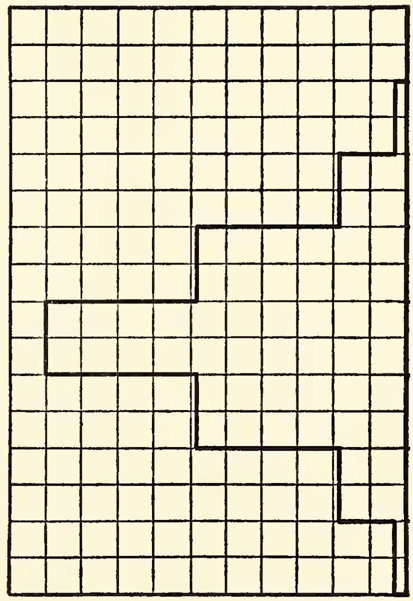


FIG. 4.—Memory span for digits, 123 adult women students. (After Thorndike, page 99)

DIAGRAM VIII

SAMPLES OF FREQUENCY DISTRIBUTIONS DRAWN FROM DIFFERENT FIELDS.

1. Elementary Principles of Probability. The Derivation and Construction of the Probability Curve

Perhaps the simplest approach to an understanding of the Normal Curve is through a consideration of the elementary facts of probability. As used in statistics, the "probability" of the occurrence of an event may be defined as the expected relative frequency of occurrence of the given event in a very large (infinite) number of observations. This expected relative frequency of occurrence may be based upon a knowledge of the conditions determining the probable occurrence, as in dice throwing or coin tossing, or upon empirical data, as in mental and social measurements.

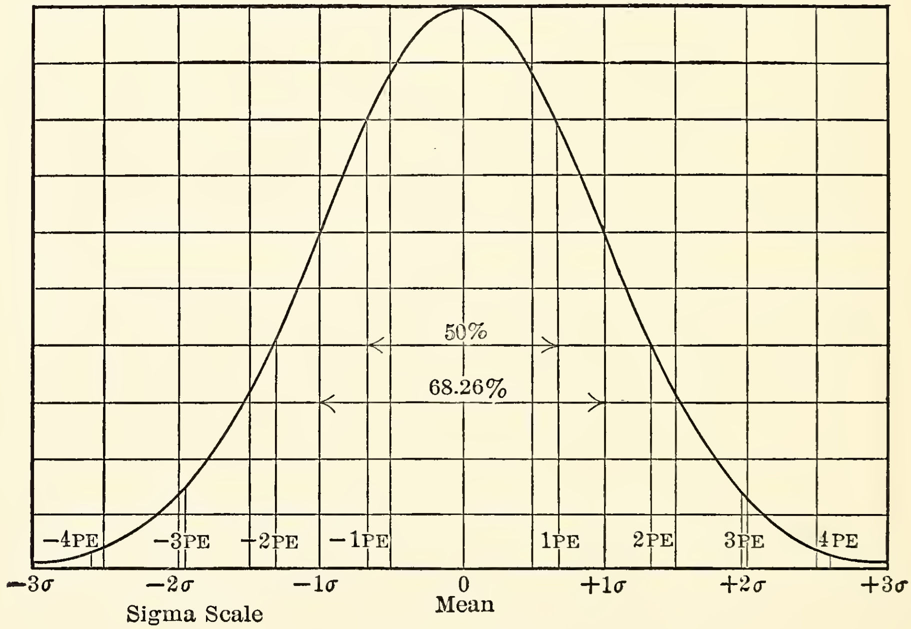


DIAGRAM IX

Normal Probability Curve.

The probability of an event may be stated most simply, perhaps, as a *ratio;* as, for example, when we say that the probability of a coin falling heads or tails is 1/2, or that of a die showing a two spot is 1/6. This ratio, called the "probability

ratio," may be defined as that fraction the numerator of which equals the expected outcome or outcomes and the denominator of which equals the total possible outcomes. Such a ratio always falls between the limits 0 (impossibility of occurrence) and 1.00 (certainty of occurrence). Thus the probability that the sky will fall is 0; that an individual now living will some day die is 1.00. Between these limits there are all possible degrees of probability expressed by the probability ratio.

Let us now apply these simple principles of probability to the specific case of what happens when we toss coins (coin tossing and dice throwing furnish simple and often-used illustrations of the laws of chance). If we toss one coin, obviously it must fall either heads (H) or tails (T) 100% of the time and a head or tail is equally probable. Expressed as a ratio, the probability of an H is 1/2; of a T, 1/2; and

$$(H+T), \text{ i.e., } \frac{1}{2}+\frac{1}{2}=1.00.$$

Again, if we toss two coins, (*a*) and (*b*), at the same time there are 4 possible arrangements which the coins may take:

(1)	(2)	(3)	(4)
a b	*a b*	*a b*	*a b*
H H	H T	T H	T T

That is, both coins (*a*) and (*b*) may fall H; (*a*) may fall H and (*b*) T; (*b*) may fall H and (*a*) T; or both coins may fall T. Expressed as a probability ratio, the chances of 2 heads are 1/4; of one head and one tail, 2×1/4 or 1/2; of 2 tails 1/4.

Let us go a step further and increase the number of coins to three. If we toss three coins, (*a*), (*b*), and (*c*) simultaneously there are 8 possible outcomes:

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
a b c	*a b c*	*a b c*	*a b c*	*a b c*	*a b c*	*a b c*	*a b c*
H H H	H H T	H T H	H T T	T H H	T H T	T T H	T T T

Expressed as a ratio, the chances of 3 heads are 1/8 (combination 1); of 2 heads and 1 tail 3/8 (combinations 2, 3, and 5);

of 1 head and 2 tails 3/8 (combinations 4, 6, and 7); and of 3 tails 1/8 (combination 8). In exactly this same way we can figure the probability of different combinations when we have 4, 5, or any number of coins.

These probable outcomes may be secured in a very much simpler way than by listing all of the various possible combinations as shown above. If there are two independent events, the probability of the occurrence or non-occurrence of each being the same (as in the probability, of a coin falling heads or tails) the "compound" probabilities may be found by the expansion of the binomial $(p+q)^2$ in which p equals the probability of its happening, q the probability of its not happening, and the exponent 2 indicates the number of events. Now if we substitute H for p, and T for q (tails = non-heads), we have for two coins $(H+T)^2$: and squaring, the binomial $(H+T)^2 = H^2+2HT+T^2$. This expansion may be written,

1	H^2	1 chance in 4 of 2 heads; *probability ratio* = 1/4
2	HT	2 chances in 4 of 1 head and 1 tail; *probability ratio* = 1/2
1	T^2	1 chance in 4 of 2 tails; *probability ratio* = 1/4
Total = 4		

Note that these results are identical with those obtained above by listing the various possible outcomes when two coins are tossed.

If we have three independent events, the expression $(p+q)^3$ becomes, for three coins, $(H+T)^3$. Expanding this binomial, we get $H^3+3H^2T+3HT^2+T^3$ which may be written,

1	H^3	1 chance in 8 of 3 heads; *probability ratio* = 1/8
3	H^2T	3 chances in 8 of 2 heads and 1 tail; *probability ratio* = 3/8
3	HT^2	3 chances in 8 of 1 head and 2 tails; *probability ratio* = 3/8
1	T^3	1 chance in 8 of 3 tails; *probability ratio* = 1/8
Total = 8		

Again these results are identical with those got by listing the various possible outcomes obtained by tossing three coins.

The binomial expansion may be applied more generally to the case in which there are any number of independent events, just so long as the probability of occurrence or non-occurrence is the same for each separate event. Thus if we toss 10 coins simultaneously, we have by analogy with the above $(p+q)^{10}$, which equals $(H+T)^{10}$, putting H for probability of a head, T for probability of a non-head (tail) and 10 for the number of coins tossed. When the expression $(H+T)^{10}$ is expanded, we have,[1]

$$H^{10}+10H^9T+45H^8T^2+120H^7T^3+210H^6T^4+252H^5T^5+210H^4T^6 +120H^3T^7+45H^2T^8+10HT^9+T^{10}$$

which may be summarized as follows:

		Probability Ratio
1 H^{10}	1 chance in 1024 of all coins falling heads. . .	$\frac{1}{1024}$
10 H^9T	10 chances in 1024 of 9 heads and 1 tail.	$\frac{10}{1024}$
45 H^8T^2	45 chances in 1024 of 8 heads and 2 tails.	$\frac{45}{1024}$
120 H^7T^3	120 chances in 1024 of 7 heads and 3 tails.	$\frac{120}{1024}$
210 H^6T^4	210 chances in 1024 of 6 heads and 4 tails.	$\frac{210}{1024}$
252 H^5T^5	252 chances in 1024 of 5 heads and 5 tails.	$\frac{252}{1024}$
210 H^4T^6	210 chances in 1024 of 4 heads and 6 tails.	$\frac{210}{1024}$
120 H^3T^7	120 chances in 1024 of 3 heads and 7 tails.	$\frac{120}{1024}$
45 H^2T^8	45 chances in 1024 of 2 heads and 8 tails.	$\frac{45}{1024}$
10 HT^9	10 chances in 1024 of 1 head and 9 tails.	$\frac{10}{1024}$
1 T^{10}	1 chance in 1024 of all coins falling tails.	$\frac{1}{1024}$
Total = 1024		

These results are represented graphically in Diagram X, by a histogram and frequency polygon plotted on the same axes. The eleven terms of the expansion have been laid off at equal distances on the X-axis, and the chances of the *occurrence* of each combination of H's and T's plotted as scores on the Y-axis. The result is a symmetrical probability curve, with the greatest concentration in the center, and the "scores" (the chances) falling away by corresponding decrements above and

[1] The reader may take this expansion on faith; or he may refer to the chapter on Binomials in any elementary Algebra.

below the central point. Diagram X represents the results which we should expect to get *theoretically* by tossing 10 coins 1024 times.

Many experiments have been made for the purpose of checking the theoretical against the actual results, by tossing coins or throwing dice a great many times. In one well-known experiment[1] 12 dice were thrown 4096 times, each

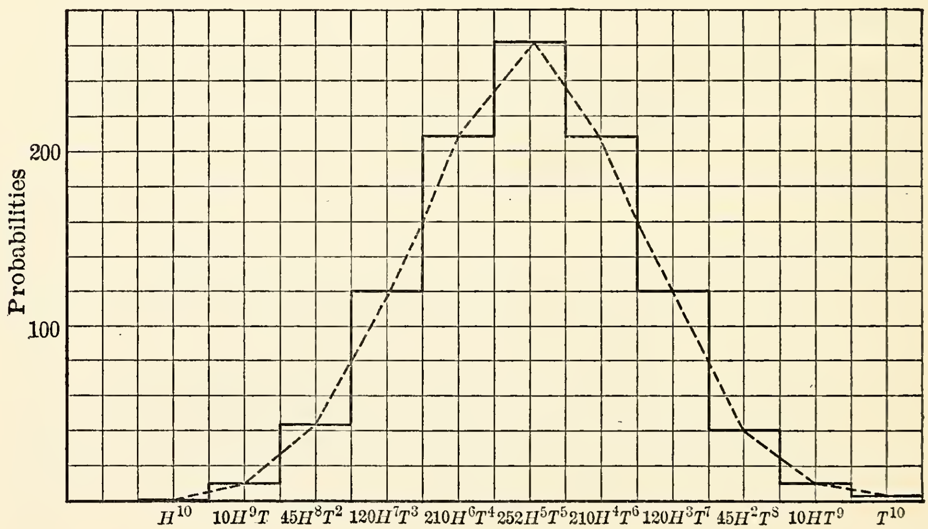


DIAGRAM X

Probability Surface Obtained from the Expansion of $(H+T)^{10}$.

4, 5, and 6 spot being taken as a "success" and each 1, 2, and 3 spot as a "failure." For example, in a throw of 3, 1, 2, 6, 4, 6, 3, 4, 1, 5, 2, 3, there would be 5 successes. The observed frequencies of the different number of successes and the theoretical results secured from the binomial expansion have been plotted on the same axes in Diagram XI. The reader will note how closely the observed frequencies check the theoretical: how close the two polygons are to being identical. If the reader should care to verify the results of Diagram XI by tossing 10 coins 1024 times, he will find his

[1] Yule G. Udny, *An Introduction to the Theory of Statistics*, 5th edition, 1919, p. 258.

empirical results closely in accord with the theoretical expectations.

2. Why the Probability Curve is Employed in Psychological Measurement

The frequency curve plotted in Diagram X from the expansion of the expression $(H+T)^{10}$ is a symmetrical 10-sided polygon. If the number of factors (e.g., coins) is increased

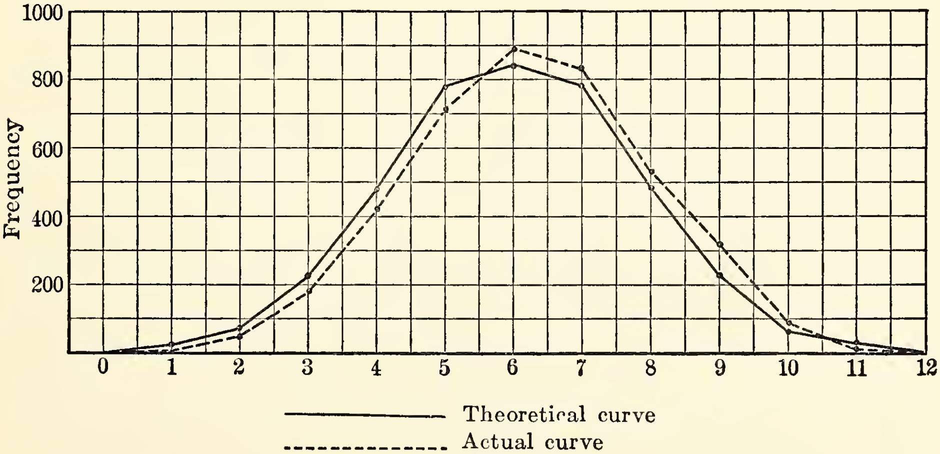


DIAGRAM XI

Comparison of Observed and Theoretical Results in Throwing 12 Dice 4096 Times. (After Yule, page 258.)

from 10 to 20, to 30, and then to 40 (the baseline extent remaining the same) the number of sides of the polygon will increase from 10 to 20, to 30, to 40. With each increase in the number of factors, the points on the curve will move more and more closely together, until finally when the number of factors becomes very large [when n in the expression $(p+q)^n$ becomes infinite] the polygon will become a perfectly smooth curve like the one in Diagram IX. The "ideal" polygon or normal curve, therefore, may be said to represent the relative frequency of occurrence of various combinations of a *very large* number of *equal*, *similar*, and *independent* factors, when the chances of the occurrence or non-occurrence of each factor is the same.

If now we compare the frequency curve in Diagram IX with the four graphs plotted from actual data obtained in measurements of height, intelligence (*IQ*), memory span, and temperature (see Diagram VIII) the similarity—as noted above—of these graphs to the normal curve is clearly evident. In other words, these distributions of variable phenomena act as though they were determined by the operation of factors which are present or absent according to the same laws which govern the combinations of coins and dice. This is found to be true of many other distributions as well; so that the general tendency of quantitative data to follow the normal probability curve is often called the "law of normal frequency." Stated briefly, this law is as follows: measurements of natural phenomena as well as measurements of mental and social traits tend to be distributed symmetrically about their central tendency in proportions which are determined by the laws of chance.

The reason why frequency distributions of variable phenomena are similar to chance distributions obtained from tossing coins or throwing dice is that the former, like the latter, are probably due very often to the operation of the laws of chance. "Chance" may be defined as the result obtained from the operation of a great many factors, none of which is dominant, or, put in another way, all of which are (relatively) similar, equal, and independent. A number of small factors, for example, determine whether a coin will fall heads or tails, or whether a die will show a 2, 3, or 6 spot: the twist of the wrist, height from which coin or die is thrown, weight or size of coin or die, kind of floor on which experiment is made, and many others.[1] In like manner a man's height, or his weight, or the shape of his head, or his intelligence, or his eye color is determined, very probably, by a large number of factors which have approximately the same influence on the final result. (Note: Should one or more of these factors have special weight the distribution will no longer be of the prob-

[1] See Jerome Harry, *Statistical Methods*, 1924, pp. 169–170.

ability type, but will be *skewed* or shifted over towards the upper or the lower end of the scale. The question of "skewness" will be considered on page 86.)

Experiments have shown that the normal probability curve serves to describe the frequency of occurrence of many variable facts with a relatively high degree of accuracy. Some of these distributions have already been shown in Diagram VIII. Important facts which give normal, or approximately normal, distributions may be classified as follows:[1]

1. Biological statistics: the proportions of male to female births for the same country or community over a period of years; the proportion of different types of plants and animals in cross-fertilization (the Mendelian ratios).

2. Anthropometrical statistics: height, weight, cephalic index, etc., for large groups of same age and sex.

3. Social and economic statistics: rates of birth, marriage, or death, under uniform conditions; wages and output of large numbers of workers under like conditions and in same occupation; labor costs, prices, etc.

4. Psychological measurements: intelligence as measured by standard tests; speed of association, perception, reaction time, etc.; educational test scores, e.g., in spelling, arithmetic, reading.

5. Errors of observation: measures of height, speed of movement, magnitudes, physical and mental traits, etc., contain errors which are as likely to cause them to lie above as below the true value Such errors follow the normal probability curve. (This topic is treated in Chapter III.)

The normal curve is often called the normal probability curve because it gives the theoretical probabilities of the occurrence of chance phenomena. It is also called the normal frequency curve because frequency distributions of actual data obtained from the measurement of many variable facts are normal. Finally, it is called the "curve of error" because when repeated measurements have been made of such variables as height,

[1] Jones D. Caradog, *A First Course in Statistics*, 1921, p. 233.

linear magnitudes, time and extent of movement, reaction, time, etc, the separate measures tend to diverge from the "true" measure (or standard) by amounts which when plotted give the characteristic probability curve (see Chapter III).

We may conclude this discussion of the normal curve with a word of caution. Despite the similarity of actual and chance distributions, the student must be careful not to draw the conclusion that because of this analogy, we can assume forthwith that mental and physical traits are always (or necessarily) due to the operation of equal, similar, and independent factors governed entirely by chance. The factors which determine, say, musical ability or intelligence are too little known to warrant the assumption, *a priori*, that they operate in the same manner, and in accordance with the same laws, as those factors which give chance distributions of coins or dice. The selection of the normal curve, rather than some other type of curve, is, after all, sufficiently justified by the fact that it does generally fit the data better. However "the theoretical justification and the empirical use of the curve are two quite different matters." [1]

3. Important Properties of the Normal Frequency Curve

In the normal frequency curve, the average, the median, and the mode all fall exactly at the midpoint of the distribution, and hence are numerically equal. This follows from the fact that the normal probability curve is perfectly symmetrical bilaterally, and in consequence all of the measures of central tendency must fall at the middle of the curve. Also in the normal curve, the measures of variability include certain constant fractional amounts of the total area of the curve as follows (see Diagram IX):

1. If the *SD* is laid off in the plus and minus directions from the mean (to right and left) along the baseline, and if perpendiculars are erected at these points, the area included

[1] Jones D. Caradog, *ibid.*, p. 233.

by the perpendiculars, the baseline, and the curve itself contains the middle 68.26% of the total area under the curve. Stated briefly, between the mean and $\pm 1\sigma$ are found the middle 2/3 (approximately) of the cases in the normal distribution.

2. If the AD is laid off in the plus and minus directions from the mean along the baseline, and if perpendiculars are erected at these points, the area included by the perpendiculars, the baseline, and the curve, contains the middle 57.5% of the total area. Put briefly, between the mean and $\pm 1AD$ will be found the middle 57.5% of the cases in the distribution.

3. If the PE is laid off in the plus and minus directions from the mean along the baseline, and if perpendiculars are erected at these points, the area included by the perpendiculars, the baseline and the curve contains the middle 50% of the area. Since the PE (equivalent to the Q in a normal distribution) equals 1/2 the distance between the 75th and 25th percentiles, in a perfectly symmetrical distribution it marks off the 25% of the area directly above and the 25% directly below the mean—the middle 50% of the measures.

Certain constant relations will be found to obtain among the measures of variability. These are easily derived from the per cents of area included by each.

1. $PE = .6745\,\sigma$
2. $PE = .8453AD$
3. $\sigma = 1.4825PE$
4. $\sigma = 1.2533AD$
5. $AD = .7979\,\sigma$
6. $AD = 1.1843PE$

The first of these relations is the only one used often enough to warrant its being memorized. From these equations it should now be evident why it was stated earlier (page 27) that the σ is always greater than the AD which is, in turn, always greater than the $Q(PE)$.

4. The Measurement of Skewness

In a frequency polygon or histogram, usually the first thing which strikes the eye is the symmetry, or—what is more often the case—the lack of symmetry in the figure. In the normal curve the mean, the median, and the mode all coincide, and there is a perfect balance or symmetry between the right and left halves of the figure. In a "skewed" distribution, on the other hand, the mean, the median, and the mode fall at different points in the distribution, and the balance (or center of gravity) is thrown to one side or the other—to right or left. The degree of displacement or skewness is measured by the formula,

$$\text{Skewness} = \frac{3(\text{mean} - \text{median})}{SD}, \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and in the normal distribution, since the mean = the median, the skewness is 0. The more nearly the distribution approaches the normal type, the closer together the mean and the median, and the less the skewness.

If we apply formula (11) to the distribution of 54 Army Alpha scores in Table I, we get $-.66$ as the measure of skewness. Distributions like this one are said to be skewed *negatively*, or to the *left*: the scores are massed at the high end of the scale (the right end), and spread out gradually at the low or left end, as shown in Diagram XII. Distributions are skewed *positively* or to the *right* when the scores are massed at the low (the left) end of the scale, and spread out gradually at the high or right end (see Diagram XIII).

Formula (11) gives the measure of skewness of the distribution of 200 cancellation scores in Table II (2) as $+.003$. This indicates a very low degree of positive skewness, and shows how very closely this distribution approaches the probability type.

There are several reasons why distributions are skewed. In the first place we should hardly expect the distribution of *IQ*'s obtained from a group of 25 eight-year old boys to be normal,

nor the distribution of *IQ*'s obtained from a special class for the dull and feebleminded, even though the latter group

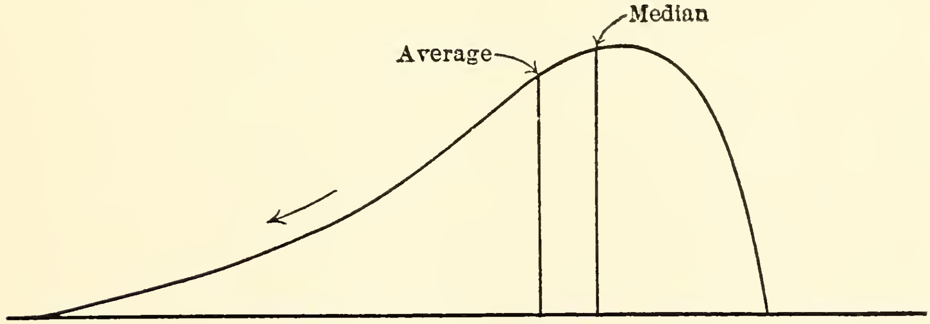


DIAGRAM XII

NEGATIVE SKEWNESS: TO THE LEFT.

were large. The small size of the group in the first case, and "special selection"[1] in the second are sufficient causes of skewness.[2] Again, technical faults in the construction

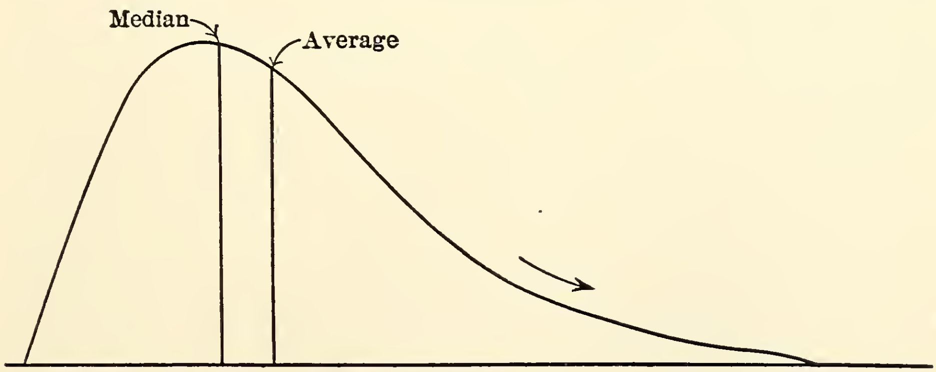


DIAGRAM XIII

POSITIVE SKEWNESS: TO THE RIGHT.

of the test, errors in scoring and the like may often produce skewness in a distribution of test scores.

In addition to these more obvious causes, skewness also

[1] A "selected" group is one which is not representative of the larger group from which it is drawn.

[2] For an illustration of skewness due to both of these causes, see the distribution of Table I.

results, oftentimes, from a real lack of "normality" in the data.[1] This condition arises when several of the factors determining a given result are dominant or prepotent and hence are present more often than chance would allow (see page 83). A simple illustration of this will be found in those distributions which result from the throwing of loaded dice. When dice of this sort are thrown, the resulting distributions will always be skewed, due to the greater "potency" of the heavier faces. Again, to take an illustration from real data, the graph representing the chances of death is considerably skewed—being higher in infancy and old age than in youth or old age—because of the difference in number and importance of the "causes of death" at certain ages.

One other illustration may be taken, this time from the field of tests. If an arithmetic test which involves only the four fundamental operations is given to 1000 eighth grade children, there will be a piling up of the scores towards the high score end of the distribution: a *negative* skewness. On the other hand, if the test contains only problems in fractions, square root, interest, etc., there will be a piling up of the scores (or at least a shift in the peak of the curve) towards the low score end of the scale: a *positive* skewness. These results may be explained in terms of the small positive and negative factors which produce the probability curve. Too easy a test excludes from operation some of the factors which make for an extension of the curve at the upper end, such as a knowledge of more advanced arithmetical relations, which the brighter children would know. Too hard a test excludes from operation factors which make for the extension of the curve at the lower end, such as a knowledge of very simple facts which would permit the answering of a few, at least, of the questions had these been included.

[1] Theoretically, there is no real reason why distributions should always be normal. Thorndike has written: "There is nothing arbitrary or mysterious about variability which makes the so-called normal type of distribution a necessity, or any more rational than any other sort, or even more to be expected on a priori grounds. Nature does not abhor irregular distributions."—*Mental and Social Measurements*, pp. 88–89.

In the one case we have a number of perfect scores, and little discrimination; in the second case a number of zero scores, and equally poor discrimination.

IV. Some Practical Applications of the Normal Curve

The entire area under any frequency curve represents the total number of frequencies in the distribution (see page 62). If we know the total area of the curve, therefore, and in addition the proportion of the total area in a given segment, it is possible to compute very simply the frequency represented by the segment. This information in regard to the normal curve is given in Tables X and XI from which the theoretical frequency of any fractional part of the probability curve may be easily obtained. Acquaintance with these tables is extremely valuable in the solution of a large number of varied problems. For this reason before considering any problems which depend for their solution on the assumption of the normal distribution, it is very desirable that the construction and use of Tables X and XI be clearly understood.

1. The Construction and Use of Tables X and XI

Table X gives the fractional parts of the total area under the normal curve found between the mean and ordinates erected at various distances from the mean, such distances measured in σ units.[1] The total area of the curve (the number of cases in the distribution) is taken arbitrarily as 10,000 because of the greater ease with which fractional parts of area may then be calculated. The first column of the table, $\frac{x}{\sigma}$, gives the distances in tenths of σ measured off on the baseline from the mean as the 0 point or origin; distances in hundredths of σ are given by the headings of the columns. To find the number of cases in a normal distribution between the mean and the ordinate erected at a distance of 1σ from

[1] Table X should be studied in conjunction with Diagram IX.

the mean, we go down the $\frac{x}{\sigma}$ column until 1.0 is reached, and in the next column under .00 take the entry opposite 1.0, viz., 3413. This figure means that there are 3413 cases in 10,000, or 34.13% of the entire area of the curve between the mean and 1σ; or put more exactly, 34.13% of the cases in the normal distribution fall within the interval bounded by the baseline, the Y-ordinate erected at the mean, the Y-ordinate erected at a distance of 1σ from the mean, and the curve itself (see Diagram IX for illustration). To find the per cent of the distribution between the mean and 1.57σ we go down the $\frac{x}{\sigma}$ column to 1.5, then across horizontally to the column headed .07 and take the entry 4418. This means that in a normal distribution, 44.18% of the entire distribution falls between the mean and 1.57σ.

Thus far we have considered only σ distances measured in the *positive* direction from the mean; that is, we have taken account only of the right half—the high score end—of the normal curve. Since the curve is bilaterally symmetrical, however, the entries in Table X may be used for σ distances measured in the *negative* (to the left) as well as the *positive* direction. Accordingly, to find the per cent of the distribution between the mean and -1.26σ, we simply take the entry 3962 in the table: the entry in the column headed .06 opposite 1.2 in the $\frac{x}{\sigma}$ column. This means that 39.62% of the cases in the distribution fall between the mean and -1.26σ. In the same way, the percentage of cases between the mean and -1.00σ is found to be 34.13; and the student will now be able to verify the statement made on page 85 that between the mean and $\pm 1.00\sigma$ are 68.26% of the cases in the normal distribution.

While theoretically the normal curve meets the baseline at infinite distances to the right and left of the mean, for practical purposes the curve may be taken to end at points

TABLE X

Fractional Parts of the Total Area (Taken as 10,000) under the Normal Probability Curve, Corresponding to Distances on the Baseline between the Mean and Successive Points Laid off from the Mean in Units of Standard Deviation.

Example: between the mean, and a point $1.3\,\sigma \left(\frac{x}{\sigma} = 1.3\right)$, is found 40.32% of the entire area under the curve.

$\frac{x}{\sigma}$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0.2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0.6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0.7	2580	2611	2642	2673	2704	2734	2764	2794	2823	2852
0.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0.9	3159	3186	3212	3238	3264	3290	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4383	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2.0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2.9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986
3.0	4986.5	4986.9	4987.4	4987.8	4988.2	4988.6	4988.9	4989.3	4989.7	4990.0
3.1	4990.3	4990.6	4991.0	4991.3	4991.6	4991.8	4992.1	4992.4	4992.6	4992.9
3.2	4993.129									
3.3	4995.166									
3.4	4996.631									
3.5	4997.674									
3.6	4998.409									
3.7	4998.922									
3.8	4999.277									
3.9	4999.519									
4.0	4999.683									
4.5	4999.966									
5.0	4999.997133									

From: *Tables for Statisticians and Biometricians*, edited by Karl Pearson, Cambridge University Press.

-3σ and $+3\sigma$ from the mean. We find from Table X, for example, that 4986.5 cases in the total 10,000 fall between the mean and 3σ; and 4986.5 cases will, of course, fall between the mean and -3σ also. Therefore, since 9973 cases in 10,000, or 99.73% of the distribution, fall within the limits set by -3σ and $+3\sigma$, by cutting off the curve at these two points we disregard only .27 of 1% of the distribution—a negligible amount, except in very large samples.

Instead of σ, the PE may be used as the unit of measurement in determining the theoretical frequencies within given intervals of the normal curve. Table XI gives the fractional parts of the total area under the normal curve found between the mean and ordinates erected at various PE distances from the mean. The table is read in exactly the same way as Table X. To find, for instance, the number of cases between the mean and $1PE$ (or more accurately the ordinate erected at that point) we go down the $\frac{x}{PE}$ column to 1.0 and in the next column under .00 read 2500. Twenty-five per cent of the cases in the distribution, therefore, lie between the mean and $1PE$. In like manner 25% of the cases lie between the mean and $-1PE$; hence, it is clear that the middle 50% of the distribution is contained between the mean and $-1PE$ and $+1PE$ measured off from the mean. This table does not read in as fine units as Table X, only tenths and .05ths PE divisions being given. If smaller divisions are desired, however, interpolation can be made.

Just as it is customary to disregard that part of the curve beyond the limits $\pm 3\sigma$, so we ordinarily disregard that part of the curve beyond the limits $\pm 4PE$. This is done because 9930 cases (4965×2) in the total 10,000 fall between the mean and $\pm 4PE$ (see Table XI). Hence, in cutting of the curve at $+4PE$ and $-4PE$, we disregard only .70 of 1% of the cases in the distribution.

There is little to choose as between Tables X and XI. The former admits of slightly easier interpolation, but the latter

is probably accurate enough, without interpolation, for most of the work done in psychological measurement.

TABLE XI

FRACTIONAL PARTS OF THE TOTAL AREA (TAKEN AS 10,000) UNDER THE NORMAL PROBABILITY CURVE, CORRESPONDING TO DISTANCES ON THE BASELINE BETWEEN THE MEAN AND SUCCESSIVE POINTS LAID OFF FROM THE MEAN IN UNITS OF *PE*.

Example: we find between the mean and a point 1.55 *PE* $\left(\frac{x}{PE}=1.55\right)$ from the mean 35.21% of the entire area under the curve.

$\frac{x}{PE}$	.00	.05	$\frac{x}{PE}$	.00	.05
0	0000	0135	3.0	4785	4802
.1	0269	0403	3.1	4817	4831
.2	0536	0670	3.2	4845	4858
.3	0802	0933	3.3	4870	4881
.4	1063	1193	3.4	4891	4900
.5	1321	1447	3.5	4909	4917
.6	1571	1695	3.6	4924	4931
.7	1816	1935	3.7	4937	4943
.8	2053	2168	3.8	4948	4953
.9	2291	2392	3.9	4957	4961
1.0	2500	2606	4.0	4965	4968
1.1	2709	2810	4.1	4971	4974
1.2	2908	3004	4.2	4977	4979
1.3	3097	3188	4.3	4981	4983
1.4	3275	3360	4.4	4985	4987
1.5	3441	3521	4.5	4988	4989
1.6	3597	3671	4.6	4990	4991
1.7	3742	3811	4.7	4992	4993
1.8	3896	3939	4.8	4994	4995
1.9	4000	4057	4.9	4995	4996
2.0	4113	4166	5.0	4996	4997
2.1	4217	4265	5.1	4997.1	4997.4
2.2	4311	4354	5.2	4997.7	4998
2.3	4396	4435	5.3	4998.2	4998.4
2.4	4472	4508	5.4	4998.6	4998.8
2.5	4541	4573	5.5	4999	4999.1
2.6	4602	4631	5.6	4999.2	4999.3
2.7	4657	4682	5.7	4999.4	4999.5
2.8	4705	4727	5.8	4999.55	4999.6
2.9	4748	4767	5.9	4999.65	4999.7

2. A Variety of Problems Solved by Means of Tables X and XI

Under this heading we shall consider a number of problems which may be solved by means of Tables X and XI, on the assumption that the distributions which they involve are normal or approximately normal. For easy reference later, each group of examples is preceded by a general statement of the problem which they are designed to illustrate.

A. To Determine the Per Cent of Cases in a Normal Distribution which Fall within Given Limits.

> *Problem* (1)—Given a normal distribution with Average $=12$, and $\sigma=4.00$. (*a*) What per cent of the cases fall between 8 and 16? (*b*) What per cent of the cases lie above 18? (*c*) below 6?

(*a*) A score of 16 is just 4 points above the mean, and a score of 8 is just 4 points below the mean. If we divide this difference of 4 points by the σ of the distribution (by 4) it is clear that 16 is 1σ above the mean and that 8 is 1σ below the mean (see Diagram XIV, Fig. I). 68.26% of the cases in a normal distribution fall between the mean and $\pm 1\sigma$ (Table X). Hence, 68.26% of the scores in the given distribution, or approximately the middle 2/3, fall between 8 and 16. This result may also be stated in terms of "chances." Since 68.26% of the cases in the distribution fall between 8 and 16, the chances are 6826 in 10,000 or 68 in 100 that any score in the distribution will be found between 8 and 16.

(*b*) A score of 18 is 6 points or 1.5σ above the mean. From Table X we find that 43.32% of the cases fall between the mean and 1.5σ. Accordingly, 6.68% of the cases (50%−43.32%) must lie *above* 18, in order to fill out the 50% of cases in the right half of the curve (see Fig. 1). Stated as "chances," there are 668 chances in 10,000 or about 7 in 100 that any future score will lie above 18.

(*c*) A score of 6 is -1.5σ from the mean. Between the

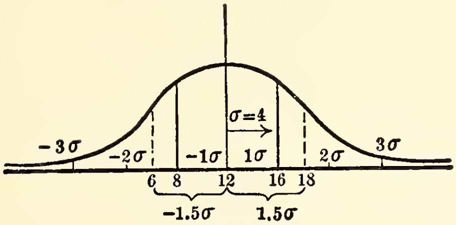


FIG. 1.

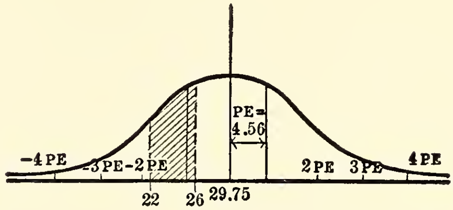


FIG. 2.

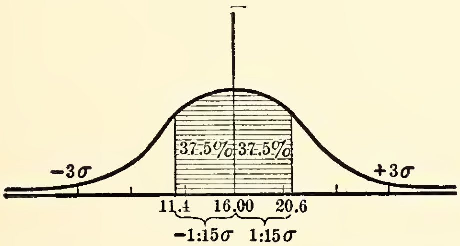


FIG. 3.

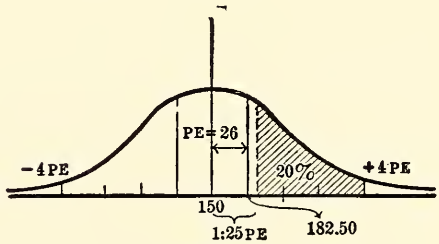


FIG. 4.

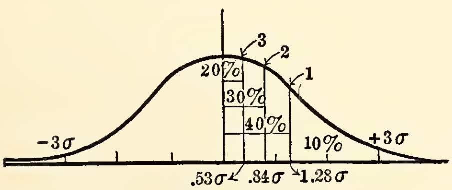


FIG. 5.

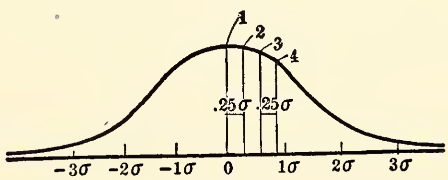


FIG. 6.

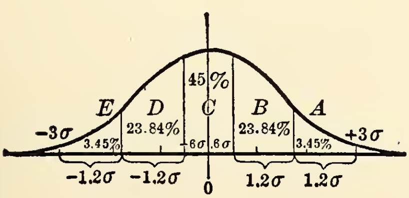


FIG. 7.

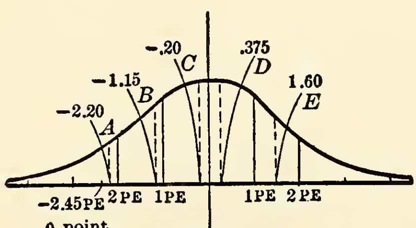


FIG. 8.

DIAGRAM XIV

ILLUSTRATING A VARIETY OF PROBLEMS SOLVED BY MEANS OF TABLES X AND XI.

mean and -1.5σ (6) are 43.32% of the cases in the entire distribution. Hence, 6.68% of the cases lie below 6—fill out the 50% below the mean—and the chances are 7 in 100 that any future score will lie below 6.

> *Problem* (2)—Given a distribution with Average$=29.75$, and $Q=4.56$. What per cent of the distribution lies between 22 and 26? What are the chances that a score will fall between 22 and 26?

In a normal distribution Q is equal to the PE. Score 22 is 7.75 points or $-1.70PE$ $\left(\text{since } \frac{7.75}{4.56}=1.70\right)$ from the mean, and score 26 is 3.75 points or $-.822PE$ from the mean (see Diagram XIV, Fig. 2). From Table XI, we find that 37.42% of the cases in a normal distribution lie between the mean and $-1.70PE$; and that 21% of the cases lie between the mean and $-.822PE$. By simple subtraction, therefore, 16.42% of the cases fall between $-1.70PE$ and $-.822PE$ or between 22 and 26. The chances are 1642 in 10,000 or 16 in 100 that a score will fall between 22 and 26.

B. To Find the Limits in Any Normal Distribution which Will Include a Given Per Cent of the Cases

> *Problem* (1)—Given a distribution with Average$=16$, and $\sigma=4$. What limits will include the middle 75% of the cases?

The middle 75% of the cases in a normal distribution must include the 37.5% just above and the 37.5% just below the mean. From Table X, we find that 3749 cases in 10,000, or 37.5% of the distribution fall between the mean and 1.15σ; and consequently, 37.5% of the distribution must fall between the mean and -1.15σ. The middle 75% of the cases, therefore, lie between the mean and $\pm1.15\sigma$; or since σ equals 4, between the mean and ±4.60 points. Adding ±4.60 to the mean (to 16), we find that the middle 75% of the scores in the given distribution lie between 20.60 and 11.40 (see Diagram XIV, Fig. 3).

Problem (2)—Given a distribution with Average$=150$, and $Q=26$. What limits will include the highest 20% of the group?

The highest 20% of a normally distributed group must have 30% of the cases between its lower limit and the mean in order to fill out the 50% of cases in the right half of the distribution (see Diagram XIV, Fig. 4). From Table XI, we find that 3004 cases in 10,000, or 30% of the distribution, fall between the mean and $1.25PE$. Since the PE of the given distribution is 26, $1.25PE$ will be 1.25×26 or 32.5 points above the mean, namely, at 182.50. The lower limit of the highest 20% of the given group, therefore, is 182.50; and the upper limit is the highest score in the distribution, whatever that may be.

C. To Determine the Relative Difficulty of Test Questions, Problems, or Other Test Items

Problem (1)—Given a test question or problem solved by 10% of a large unselected group; a second problem solved by 20% of the group; and a third, solved by 30% of the group. Assuming that the capacity measured by the test problems is distributed "normally" what is the relative difficulty of questions 1, 2, and 3?

Our first task is to find for question 1 a position in the distribution, above which are 10% (the per cent passed) and below which are 90% (the per cent failed) of the entire group. The highest 10% in a normally distributed group has 40% of the cases between its lower limit and the mean ($50\%-10\%=40\%$, see Diagram XIV, Fig. 5), and from Table X we find that 39.97%, i.e., 40%, of a normal frequency distribution falls between 1.28σ and the mean. Hence, question 1 falls at a point on the baseline of the curve whose abscissa is 1.28σ from the mean; and accordingly 1.28σ may be taken as its difficulty value.

In the same way, question 2, passed by 20% of the group, falls at a point in the distribution 30% above the mean

(50%−20%=30%, see Fig. 5). From Table X we find that 29.95%, i.e., 30%, of the group falls between the mean and $.84\sigma$; hence question 2 has a difficulty value of $.84\sigma$. In like manner question 3, which falls at a point in the distribution 20% above the mean has a difficulty value of $.53\sigma$, since 20.19% of the distribution lies between the mean and $.53\sigma$.

To summarize our results:

Question	Passed by	σ value	σ difference
1	10%	1.28	—
2	20%	.84	.44
3	30%	.53	.31

The σ difference in difficulty between 2 and 3 is .31, roughly only 3/4 of the σ difference in difficulty between 1 and 2 (.44) in spite of the fact that the per cent difference is the same in the two cases. On the assumption that ability follows the normal frequency distribution, therefore, it is evident that the σ and not the per cent difference gives the real index of differences in difficulty.

Problem (2)—Given three test items, No. 1, No. 2, and No. 3, passed by 50%, 40%, and 30%, respectively, of a large group. What per cent of the same group must pass test item No. 4, in order for it to be as much more difficult than No. 3, as No. 2 is more difficult than No. 1?

A question or problem which is "passed" by 50% of a group is, of course, "failed" by 50% also, and accordingly, such a problem falls exactly in the middle of normal distribution of difficulty. Test item 1, therefore, has a σ value of 0; it falls just on the mean (see Diagram XIV, Fig. 6). Test item 2 lies at a point in the distribution 10% above the mean, as 40% of the group passed, and 60% failed this problem. Accordingly, the σ value of this item is .25, since from Table X, we find that 9.87%—roughly 10%—of the cases lie between the mean and $.25\sigma$. Test item 3, passed by 30% of the group, lies at a point 20% above the mean, and this item, therefore, has a difficulty value of $.52\sigma$ as 19.85% (20%) of the normal distribution lies between the mean and $.52\sigma$.

Now since item 2 is $.25\sigma$ further along on the difficulty scale (towards the high score end of the curve) than item 1, it is clear that item 4 must be $.25\sigma$ above item 3, if it is to be as much harder than 3 as 2 is harder than 1. Item 4, therefore, must have a value of $.52\sigma + .25\sigma$ or $.77\sigma$; and from Table X, we find that 27.94% of the group fall between the mean and this point. This means that $50\% - 28\%$ or 22% of the group pass item 4. To summarize by a table:

Test Item	Passed by	Difficulty Value (σ)	σ difference
1	50%	.00	—
2	40%	.25	.25
3	30%	.52	—
4	22%	.77	.25

A problem or test item must be passed by 22% of the group, therefore, in order for it to be as much more difficult than an item passed by 30%, as an item passed by 40% is more difficult than one passed by 50%. Note again that per cent differences are not reliable indices of differences in difficulty when the capacity measured is taken to be distributed normally.

D. To Separate a Given Group into Sub-Groups According to Capacity, When the Capacity is Normally Distributed

Problem (1)—Suppose that we have measured 100 college men on a certain test. We wish to classify our group into 5 sub-groups A, B, C, D, and E, according to ability, the *range of ability* to be equal in each sub-group. Assuming that the capacity measured by the test is distributed normally, or approximately so, and that the group is relatively unselected, how many men should be placed in groups A, B, C, D, and E, respectively?

Let us first represent the positions of the five sub-groups graphically on the normal curve as shown in Diagram XIV, Fig. 7. If the baseline of the curve is taken to extend from -3σ to $+3\sigma$, that is, over a range of 6σ, dividing this range by 5, we get 1.2σ as the baseline extent to be allotted to each group.

These five intervals may be laid off on the baseline as shown in the figure, and perpendiculars drawn to demarcate the various sub-groups. It is clear that group A covers the upper 1.2σ; group B, the next 1.2σ; that group C lies $.6\sigma$ to the right and $.6\sigma$ to the left of the mean; and that groups D and E occupy the same relative positions on the left half of the curve, as B and A occupy on the right half.

Now to find what per cent of the whole group falls within the A group, we must find what per cent of a normal distribution lies between 3σ (the upper limit of the A group) and 1.8σ (the lower limit of the A group) (see Fig. 7). From Table X we know that 49.86% of a normal distribution falls between the mean and 3σ; and that 46.41% falls between the mean and 1.8σ. Hence 3.45% of the total area under the normal curve (49.86%−46.41%) falls between 3σ and 1.8σ, and, accordingly, group A comprises 3.45% of the whole group.

The per cents in the other groups are found in exactly the same way. Thus, 46.41% of the normal curve falls between the mean and 1.8σ (upper limit of group B) and 22.57% falls between the mean and $.6\sigma$ (lower limit of the same group). Subtracting, 46.41%−22.57% or 23.84% of our whole group evidently belongs in sub-group B. Group C lies $.6\sigma$ above and $.6\sigma$ below the mean. Between the mean and $.6\sigma$ is contained 22.57% of a normal distribution, and the same per cent is contained between the mean and $-.6\sigma$. Group C, then, includes 45% (22.57%×2) of the whole group. Finally, group D which falls between $-.6\sigma$ and -1.8σ contains exactly the same percentage of the total as group B; and group E which falls between -1.8σ and -3σ contains the same per cent as group A. The percentage (and number) of men in each group is given in the following summary:

Group	A	B	C	D	E
Per cent of total in each group........	3.5	23.8	45	23.8	3.5
Number in each group (100 men in all)...	4 or 3	24	45	24	4 or 3

On the assumption that the capacity measured follows the normal probability curve, therefore, only 4 men in the group of 100 should be placed in group A—call the marked ability group; 24 in group B, the high average ability group; 45 in group C, the average ability group; 24 in group D, the low average ability group; and 4 in group E, the very low or stupid group.

The above procedure may be used in determining how many individuals in a large class should get grades of say, A, B, C, D, E, or it may be employed for any number of grade-groups. The assumption must be made, however, that the subject in which the individuals are being graded follows the normal curve.

3. The Arrangement of Problems or Other Test Items into a Scale in which the Difficulty of Each Item is Known with Reference to Each Other Item as Well as Some Selected Zero Point

One of the important tasks which confronts the worker with tests is the construction of scales which shall contain problems or questions graded in difficulty from very easy to very hard by known steps or intervals. Given a set of problems or test items, if we know what per cent of a large group (selected from among those for whom the test is intended) pass or fail each problem, it is a comparatively easy matter to arrange the problems in a rough order of difficulty. Such an arrangement, however, constitutes a very crude scale, as we know very little about the relative difficulty of the separate problems (see page 98) and next to nothing about the range of ability tested.

For this reason in most scaled tests—if we can assume a normal or approximately normal distribution in the capacity tested—the unit of measurement is taken as the σ or the PE. By so doing we are able not only to arrange the test items in a simple order of difficulty, but to "set" or space them at definite points along a scale of difficulty—along the baseline of the normal curve. On such a scale the distance from one item to another,

or from any given item to the selected zero point is known as definitely as the distance between two divisions on a foot rule. To illustrate concretely how a scale of this sort is made, let us suppose that we wish to construct a scaled test for measuring "reasoning ability" (e.g., by means of syllogisms) in 12 year olds; or an addition scale for Grade IV; or a scale for testing sentence memory in 8 year olds. The steps involved may be outlined as follows:

(1) First it is necessary to compile a large number of problems or other test items which vary in difficulty from very easy to very hard, and which are fairly representative of the field covered by the test.

(2) These problems are then given to as large a random sample as possible from among those for whom the scale is intended.

(3) The per cent of the group which solves each problem correctly is next computed. This allows duplicates and problems too easy or too hard or those which for one reason or another are unsatisfactory to be discarded. It also permits the arrangement of the problems selected for the scale into an order of difficulty. A problem solved correctly by 90% of the group is obviously easier than one solved correctly by 75%; while the second problem is, in turn, clearly less difficult than one solved correctly by 50%. The larger the per cent passing the lower the position of the problem on the difficulty scale.

(4) By means of Table XI each per cent correct found in (3) may now be converted into a PE (or σ) [1] distance above or below the mean. The procedure here is as follows. An item solved correctly by 40% of the group is 10% or $.375PE$ above the mean. In like manner, an item solved correctly by 78% of the group is 28% (78%−50%) or $1.15PE$ below the mean. We may tabulate the results for five items selected at random as follows (see Diagram XIV, Fig. 8):

[1] The procedure is identical when σ is employed instead of the PE.

Problem	A	B	C	D	E
Per cent solving	93	78	55	40	14
Distance from mean in percentage terms	−43	−28	−5	10	36
Distance from the mean in *PE* terms	−2.20	−1.15	−.20	.375	1.60

Note that Problem A is solved by 93% of the group, i.e., by the upper 50% (the right half of the curve) plus the 43% to the left of the mean. Hence it is −2.20*PE* to the left of the mean. In like manner, the percentage distance from the mean measured to the right or left—plus or minus—for each problem may be found by simply subtracting the per cent passing from 50%. From these percents, the *PE* distance of the problem from the mean can be read from Table XIV, as shown above.

(5) With the *PE* distance of each problem above or below the mean established, the *PE* distance of each problem from the "zero point" of ability in the test may be calculated. This zero point is located in the following way. Suppose that 5% of the whole group failed to solve a single problem correctly. This puts the point of zero ability 45% of the distribution below the mean or at a point −2.45*PE* from the mean.[1] The *PE* distance of each problem in the scale may now be found from this arbitrary zero point. To illustrate with the five problems above:

Problem	A	B	C	D	E
PE distance from mean	−2.20	−1.15	−.20	.375	1.60
PE distance from assumed zero, i.e., −2.45*PE*	.25	1.30	2.25	2.83	4.05

The simplest way to find the *PE* distances from the given zero point is to subtract, algebraically, the distance of the zero point below the mean, from the *PE* distance of each problem from the mean. Problem A, for example, is −2.20—(−2.45) or .25*PE* from the zero point; while problem *E* is 1.60—(−2.45) or 4.05*PE* from the zero point. The *PE* value of each of the other

[1] Note that this point is not a true zero unless the problems range down to zero difficulty. It serves, however, as a convenient reference point for the group for whom the test is intended.

problems as measured from the given zero point is found in the same way.

When the *PE* value from zero of each of the problems has been determined, the difficulty of each problem with respect to every other problem as well as to zero is known and the scale is finished.

It is evident, of course, that a scale of this sort will not usually have equal difficulty intervals or "steps" from easy to hard. However, this fact, while inconvenient, does not necessarily invalidate the usefulness of the scale as a measuring instrument. In lieu of a rule, one might use a stick on which marks had been set at 2, 3.7, 4.8, etc., inches with a fair degree of accuracy. Nevertheless linear measurements are certainly more easily obtained with a rule, and in like manner scores are more easily obtained when the scale has equal steps than when the steps are unequal. For this reason among others, scale makers have tried as far as possible to have the steps on their scales approximately equal. One method of doing this is to eliminate from the scale as first constructed, certain "odd" problems, and retain only those which fall at points approximately the same distance apart. Another plan is to try out a new set of problems, and from among these select problems which will fill in the gaps in the scale; or to change the wording or scoring of a problem in such a way as to shift it up or down on the scale of difficulty.

A good example of the first method of securing equal steps on the scale is given by the Woody Arithmetic Scales, Series B.[1] These scales represent a selection of certain problems from the longer Series A (scales constructed by the method outlined above) and contain problems which are progressively more difficult by approximately equal steps. The problems in Series A are not spaced at equal points on a difficulty scale. In the Addition Scale, for example, problem No. 1 has a difficulty value of 1.23*PE* as measured from the arbitrary zero

[1] Woody, Clifford: *Measurements of Some Achievements in Arithmetic.* Teachers College, Columbia University, 1916.

$-2.425PE$;[1] problem No. 2 has a difficulty value of $1.40PE$, and problem No. 3 a difficulty value of $2.50PE$.

TABLE XII

Difficulty Values (*PE*) of the Problems in the Woody Arithmetic Scale (Addition), Series A and B

Problem No.	Series A, *PE* Value	Series B, *PE* Value	*PE* Differences (Series B)
1	1.23	1.23	
2	1.40	1.40	.17
3	2.50	2.50	1.10
4	2.61		
5	2.83	2.83	.33
6	3.21		
7	3.26	3.26	.43
8	3.35		
9	3.63		
10	3.78	3.78	.52
11	3.92		
12	4.18		
13	4.19	4.19	.41
14	4.85	4.85	.66
15	4.97		
16	5.52	5.52	.67
17	5.59		
18	5.73		
19	5.75	5.75	.23
20	6.10	6.10	.35
21	6.44	6.44	.34
22	6.79	6.79	.35
23	7.11	7.11	.32
24	7.43	7.43	.32
25	7.47		
26	7.61		
27	7.62		
28	7.67		
29	7.71	7.71	.28
30	7.71		
31	7.97		
32	8.04		
33	8.18	8.18	.47
34	8.22		
35	8.58		
36	8.67	8.67	.49
37	8.67		
38	9.19	9.19	.52

[1] The arbitrary zero point on the Woody addition scale is $-2.425PE$ below the median of Grade II.

The number and the *PE* value of the other problems in Series A (Addition) and the problems which have been selected from this series to make up Series B are shown in Table XII. Each problem in Series A, as noted above, is expressed in terms of its *PE* distance from the arbitrary zero point $-2.425PE$ below the second grade median. The extremely high *PE* values of the problems in the upper half of the scale result from the fact that the scale is intended for the elementary grades from II to VIII inclusive, and hence the more difficult problems fall entirely out of the range of second grade ability. Note that except in a very few cases, the problems in Series B appear as a graded series from easy to hard in which the steps from problem to problem are fairly well equalized. The score on this scale is simply the number of problems solved correctly—the distance which one progresses up the scale—just as a child's height is so many feet and inches on a scale of height.

On a scale which has equal steps, we know that the increase from say point 10 to 12 is the same as the increase from 12 to 14, and 1/2 the increase from 14 to 15. Moreover, we may say that the child who works 8 problems is as far ahead of the child who works 4, as the second child is ahead of one who cannot work a single problem. We must be extremely careful not to interpret one measure of capacity on such a scale as "so many times" another measure, however. Unlike measures of height or weight which are measured from absolute zeros, the measures given by a scale of performance are taken from some arbitrary zero point selected by the experimenter. So while we may say that a man 72 inches in height is twice as tall as a child who is only 36 inches in height, we cannot, by analogy, say that a child who scores 5 on an addition test has doubled his ability when he is able to score 10, *unless* the measures in the test have been taken from the absolute zero point of " just no ability at all " in addition.

The method of constructing a scale outlined above may be used with any group, grade, or class. When the scale is designed for use with more than one group, e.g., for the whole

elementary school, an extension of the method given is often used. In brief, this is as follows:

(1) The *PE* value of each problem is determined for each grade separately, as shown above, by computing the per cent who pass each problem.

(2) The *PE* distances between the different grade medians are then computed. This is done by finding the per cent of the pupils in each grade who have scores larger than the median score of the next grade. These per cents, when turned into *PE* values by means of Table XI, give the *PE* distances between adjoining grade medians.

(3) Knowing the *PE* distances between the grade medians, we may now convert the *PE* distance of each problem from a given grade median into a *PE* distance from some common zero point. The different *PE* values of each problem as determined for the various grades are averaged to give the final scale value [1]—the distance from the common zero point.

A shorter method than the one described may also be used. This is to compute the *PE* value of a problem once for all from the per cent of a large sample—drawn from the entire group—who pass the problem. This plan is practically identical with that which we have already described on page 102. It assumes that the capacity which the scale is designed to measure is distributed normally throughout the entire group. While probably not as exact as the more elaborate method, it has the advantage of simplicity and straightforwardness.

4. The Conversion of Judgments by Relative Position—or Relative Merit—into σ or *PE* Positions on a Scale

The preceding paragraphs have dealt with the construction of performance scales built up on the principle that the per cent passing (or failing) a given problem is the best index of the difficulty of that problem. It sometimes happens, however,

[1] A method of weighting the *PE* values of a problem in averaging the results from the different grades is described by Woody in his "*Measurements of Some Achievements in Arithmetic.*"

that the ability to be measured is of such a nature that performance in it cannot be scored simply as correct or incorrect, but must be determined by a comparison with other performances of a like sort. This leads to the construction of product scales. Handwriting scales, composition scales, drawing scales are examples of instruments in which the quality of the product is measured, and not its presence or absence in terms of a per cent or number correct. For example an individual's handwriting is rated for merit by comparing it with " standard " specimens of handwriting the quality of which is known.

Quality scales are constructed on the assumption that equally often noticed differences—in merit or excellence—are equal. The first step is to secure a large number of samples of the thing to be measured, e.g., specimens of handwriting or composition, ranging from very poor to excellent. The next step is to have a large number of presumably able judges arrange these specimens in order of merit, in this way comparing each specimen with each other one. The number of times each specimen is ranked above each other one is now reduced to percentage terms, and this per cent is expressed as a *PE* difference between the two specimens. The *PE* difference determined, specimens selected for the scale may be expressed as so many *PE above* some arbitrary zero point. We may take specimens 8 and 9 on the Hillegas Composition Scale [1] as an illustration of the method. Hillegas had each of 202 judges arrange a number of English compositions in order of merit. An artificial composition was selected as being of zero merit, and given the value 0 on the scale. Of the 202 judges, 136 or 67.5% ranked 9 as better than 8. From Table XI, we know that a percentage difference of 67.5% indicates a *PE* difference of .66*PE*, and this value, therefore, expresses the amount by which 9 is better than 8. The value of 8 had already been found to be 7.72*PE* above the 0 point on the scale. Hence 9 is 7.72+.66 or 8.38*PE* above the zero compo-

[1] Hillegas, Milo B. *A Scale for the Measurement of Quality in English Composition by Young People.* Teachers College, Columbia University, 1912.

sition. The values of the other compositions on the Hillegas Scale as measured in PE values from zero, the differences determined in terms of relative merit, are 0, 1.83, 2.60, 3.69, 4.74, 5.85, 6.75, 7.72, 8.38, 9.37. Note that the steps on this scale are fairly regular, being approximately $1PE$ apart.

5. The Scaling of Total Scores on a Test

Before concluding this brief review of the methods of constructing scales, we should mention several methods used for scaling total scores on a test. The distinction between these methods and those we have outlined is that in the latter, instead of scaling each separate element on the test for difficulty—except possibly to secure an approximate order of difficulty—we simply determine the difficulty value attained as a result of doing correctly a certain number of test elements. In other words the score depends on total number of questions answered or problems worked, and the difficulty value of individual problems is not considered as in (**3**) and (**4**) above. The three methods [1] proposed for scaling total scores give, respectively, (*a*) a percentile scale, (*b*) an age scale, and (*c*) a T-scale.

(*a*) We have already learned how to locate the percentile values in a distribution of scores (pages 45–46). In a percentile scale a child making a certain score (total number correct) on a test is given a percentile rating of 20, 30, 70, etc., according to his position in the distribution. The percentile method assumes that the difference between a percentile of say 10 and 20 is the same as the difference between a percentile of 40 and 50: that percentile differences are equal throughout the scale. There is considerable reason to doubt this assumption of equal units on the percentile scale, however; and for this reason while practically very useful, the percentile scale is not entirely sound theoretically.

(*b*) In the age scale, the mean number of points scored, on the test by unselected 7 year olds is scored 7, the mean number of points scored by unselected 9 year olds is scored 9, and

[1] See McCall, W. M. *How to Experiment in Education*, 1923, p. 95ff.

so on for other age groups. Scores which fall between age groups are evaluated by interpolation. The age scale is widely used, and is easily interpreted. The chief drawback to its use seems to be the difficulty of getting unselected samples for determining the norms of the low and high age groups. Many very young children are not in the schools, while many of the older ones for one reason or another have been eliminated. As a result, age scales are only strictly accurate between very narrow ranges of ability.

(*c*) Recently McCall has suggested a method of scaling total scores, the *T*-scale, which eliminates many of the defects of both the percentile and the age scale methods. In this method, scores are based on the σ of the distribution of scores made by unselected 12 year olds. *T*-scores range from 0 to 100. The zero point on the scale is taken at 5σ below the mean and the 100 point at 5σ above the mean. The unit of measure, or one "*T*" is .1 of the σ of the distribution of unselected 12 year olds. The mean *T*-score, therefore, is 50 and each 10 points above or below this point represent 1σ of the 12 year old distribution. In actual practice *T*-scores will be found to range generally between 15 and 85. A person who stands at the mean of 12 year olds on a given test has a *T*-score of 50; one who stands 1σ above the mean, a *T*-score of 60; and one who stands 1σ below the mean of 12 year olds a *T*-score of 40.[1]

The construction of the *T*-scale has been described in great detail by McCall in Chapter X of his *How to Measure in Education*, and in consequence only the most important advantages of the scale need be considered here.[2] In the first place, the scale covers a wide range of ability which may be extended if necessary. Secondly, all *T*-scores are expressed in terms of the same unit and with respect to the same zero point and are equal throughout the scale. Accordingly, scores from different tests are directly comparable and may

[1] For an example, see the Thorndike-McCall Reading Scales, published by Teachers College, Columbia University.

[2] For a complete discussion of the advantages of the T-Scale over the age and percentile scales, see McCall, *How to Experiment in Education*, 1923, 94ff.

be combined by simple addition. Finally, a score of a given size will always have the same meaning when referred to the mean of unselected 12 year olds which remains at 50.

V. The Transmutation of Measures by Relative Position (in Order of Merit) into Measures in Units of Amount

It is often very desirable, especially in the calculation of coefficients of correlation, to be able to transmute measures arranged in order of merit into measures in units of amount or "scores" on some linear scale. This can easily be accomplished by means of tables, provided we can assume "normality" in the trait for which the ranking has been made. To take an example, let us suppose that we have 15 salesmen ranked in order of merit for selling efficiency, the most efficient ranked No. 1, the least efficient ranked No. 15. Now if we are justified in assuming that selling efficiency follows the normal probability curve, we can—with the aid of Table XIII—assign to each man a "selling score" on a scale of 10 or 100 points which will very probably represent his capacity as a salesman much better than a rank of 2, 6, or 14. The problem may be stated as follows:

> *Problem* (1)—Given 15 salesmen ranked in order of merit by their sales-manager, transmute these rankings into scores on a scale of 10 points.

The procedure is as follows: First by means of a simple formula,

$$\text{Per cent position} = \frac{100(R - .5)}{N},^{1} \quad . \quad . \quad . \quad (12)$$

in which R is the rank of the individual in the series, and N the number ranked, we determine the "per cent position" of each man. Next, from Table XIII we read off the score on a scale of 10 points. Thus Salesman A who ranks No. 1 (see the

[1] This formula and the method built around it were devised by Professor Clark Hull. See Hull, *The Computation of the Pearson r from Ranked Data*, Journal of Applied Psychology, 1922, 6, 385.

table below) has a per cent position of $\frac{100(1-.5)}{15}$ or 3.34, and his score from Table XIII is 8.5 (finer interpolation unnecessary). In like manner, Salesman B who ranks No. 2 has a per cent position of $\frac{100(2-.5)}{15}$ or 10, and his score, accordingly, is 7.5. The scores of the others, found in exactly the same way, are given in the following table:

Salesmen	Rank	Per cent Position	Score (Scale 10)
A	1	3.34	8.5
B	2	10.00	7.5
C	3	16.67	6.9
D	4	23.34	6.4
E	5	30.00	6.0
F	6	36.67	5.7
G	7	43.34	5.3
H	8	50.00	5.0
I	9	56.67	4.7
J	10	63.34	4.3
K	11	70.00	4.0
L	12	76.67	3.6
M	13	83.34	3.1
N	14	90.00	2.5
O	15	96.67	1.5

On several previous occasions, it has been pointed out that the assumption of normality in a trait or capacity implies that differences at the extremes of capacity are relatively much greater than the same differences around the average or mean. This is clearly brought out in the table above; for while all differences in the order of merit series equal 1, the differences between the transmuted scores vary considerably, being greatest at the ends of the series, and smallest in the middle. The difference between A and B, for example, or between N and O, is three times as great as the difference between G and H. Stated differently, we may say that it is three times as easy to move from H to G (from 8th to 7th place) as from B to A (from 2nd to 1st place).

TABLE XIII

[From Hull, Journal of Applied Psychology, 1922]

THE TRANSMUTATION OF AN ORDER OF MERIT INTO UNITS OF AMOUNT OR "SCORES."

Let R represent the rank in the Order of Merit, and N the number ranked. Then from the formula, Per cent position $= \frac{100(R-.5)}{N}$, find the per cent position, and from it the score.

Example: If $N=25$, and $R=3$, Per cent position $= \frac{100(3-.5)}{25}$ or 10.00, and from the table the score is 7.5.

Per cent	Score	Per cent	Score	Per cent	Score
.09	9.9	22.32	6.5	83.31	3.1
.20	9.8	23.88	6.4	84.56	3.0
.32	9.7	25.48	6.3	85.75	2.9
.45	9.6	27.15	6.2	86.89	2.8
.61	9.5	28.86	6.1	87.96	2.7
.78	9.4	30.61	6.0	88.97	2.6
.97	9.3	32.42	5.9	89.94	2.5
1.18	9.2	34.25	5.8	90.83	2.4
1.42	9.1	36.15	5.7	91.67	2.3
1.68	9.0	38.06	5.6	92.45	2.2
1.96	8.9	40.01	5.5	93.19	2.1
2.28	8.8	41.97	5.4	93.86	2.0
2.63	8.7	43.97	5.3	94.49	1.9
3.01	8.6	45.97	5.2	95.08	1.8
3.43	8.5	47.98	5.1	95.62	1.7
3.89	8.4	50.00	5.0	96.11	1.6
4.38	8.3	52.02	4.9	96.57	1.5
4.92	8.2	54.03	4.8	96.99	1.4
5.51	8.1	56.03	4.7	97.37	1.3
6.14	8.0	58.03	4.6	97.72	1.2
6.81	7.9	59.99	4.5	98.04	1.1
7.55	7.8	61.94	4.4	98.32	1.0
8.33	7.7	63.85	4.3	98.58	.9
9.17	7.6	65.75	4.2	98.82	.8
10.06	7.5	67.48	4.1	99.03	.7
11.03	7.4	69.39	4.0	99.22	.6
12.04	7.3	71.14	3.9	99.39	.5
13.11	7.2	72.85	3.8	99.55	.4
14.25	7.1	74.52	3.7	99.68	.3
15.44	7.0	76.12	3.6	99.80	.2
16.69	6.9	77.68	3.5	99.91	.1
18.01	6.8	79.17	3.4	100.00	0
19.39	6.7	80.61	3.3		
20.93	6.6	81.99	3.2		

Another use to which Table XIII may be put is in the combining of incomplete order of merit rankings. To illustrate with a problem:

Problem 2—Given six persons, A, B, C, D, E, and F, to be ranked for honesty by three judges. Judge 1 knows all six well enough to rank them; Judge 2 knows only three well enough to rank them; and Judge 3 knows four well enough to rank them. Can we obtain a fair order of merit for all six persons by combining these three sets of rankings, two of which are incomplete?

We may tabulate the data as follows:

Persons	A	B	C	D	E	F
Judge 1's ranking......	1	2	3	4	5	6
Judge 2's ranking......		2		1		3
Judge 3's ranking......	2		1		3	4

Now assuming that honesty is "normally distributed" it seems fair that A should get more credit for ranking first in a list of six than D for ranking first in a list of three, or C for ranking first in a list of four. In the order of merit rankings, all three are given the same rank. But when we assign scores to each person in accordance with his position in the list by means of formula (12) and Table XIII, A gets 77 for his first place, D gets 69 for his, and C gets 72 for his (see table below).[1]

Persons	A	B	C	D	E	F
Judge 1's ranking...............	1	2	3	4	5	6
Score.................	77	63	54	46	37	23
Judge 2's ranking...............	..	2	..	1	..	3
Score.................	..	50	..	69	..	33
Judge 3's ranking...............	2	..	1	..	3	4
Score.................	55	..	72	..	43	28
Sum of scores...................	132	113	126	115	80	84
Average score...................	66	57	63	58	40	28
Order of Merit..................	1	4	2	3	5	6

[1] It is somewhat doubtful whether it is usually worth the trouble to transmute orders of merit into scores as shown above and then combine them so as to get a weighted order (see Garrett, H. E., *An Empirical Study of the Various Methods of Combining Incomplete Order of Merit Ratings*, Journal of Educational Psychology, 1924, XV, pp. 157–171). If it is deemed desirable to weight ratings, however, the method given will prove useful.

The other ratings are transmuted in the manner shown above. All of the scores are then combined and averaged to give the final *weighted* order of merit as shown in the table.

With formula (12) and Table XIII it is possible to transmute any set of ranks into scores on the assumption of a normal distribution in the trait for which the ranking is made. This is very useful in the case of those traits which are not easily measured by ordinary methods, but for which individuals may be arranged in an order of merit, as for example athletic ability, personality, beauty, etc. It is also valuable in correlation when a set of ranks is the only available "criterion" for a given ability while the "independent" tests are scored in ordinary test units.[1] Transmuted scores may be combined, or averaged, like other test scores.

A word of explanation may be said in regard to the construction of Table XIII. This table was derived from a table of the theoretical frequencies of the normal frequency distribution in which the curve was taken to end at $\pm 2.5\sigma$. The baseline of the curve is 5σ, therefore, and may conveniently be subdivided into 100 parts, each $.05\sigma$. The first $.05\sigma$ from the upper extreme limit of the curve takes in .09% of the distribution and is scored 9.9 (or 99 on a scale of 100). The next $.05\sigma$ ($.10\sigma$ from the upper end of the curve) takes in .20% of the entire distribution and is scored 9.8, or 98, and so on. In each case, the percent position gives the fractional part of the normal distribution which lies to the right of the given σ value on the baseline. The σ values determine the score.

PROBLEMS

1. (*a*) Plot both distributions given in example (2), page 56 as frequency polygons and histograms. For comparative purposes plot the frequency polygon and the histogram for each distribution with respect to the same coordinate axes: on the same diagram.

   (*b*) Calculate a measure of skewness for both distributions.

[1] The definition of a criterion and its value in determining the validity of one or more tests is discussed at length in Chapters V and VI.

2. Plot distribution A, example (2), page 56, as an ogive. Compare the percentiles obtained from the graph with the calculated values.

3. Assuming that trait X is completely determined by 6 factors—all equal in value, similar, and independent, and each as likely to be present as absent—plot the distribution which one would most probably get from the measurement of trait X in an unselected group of 1000 people.

4. In a random sample of 1000 cases, Average $=14.4$, and $\sigma=2.5$.
   (*a*) What per cent of the cases lie between 12 and 16?
   (*b*) What are the chances that any future case will be above 18?
   (*c*) What are the chances that any future case will be below 8?

5. In an approximately normal distribution of 100 cases, Average $=29.74$, $Q(PE)=3.18$.
   (*a*) What per cent of the cases lie between 24 and 25?
   (*b*) What limits include the middle 60% of the cases?
   (*c*) What limits include the lowest 5% of the cases?

6. In a certain test the 7th grade median is 28, with a Q of 4.8; and the 8th grade median is 31.6, with a Q of 4.0. What per cent of the 7th grade is above the median of the 8th grade?

7. A group of 12 year olds, two years ago, had a reading ability expressed by an average of 40, and a σ of 3.6; and a composition ability expressed by an average of 62, and a σ of 9.6. Today the group has gained 12 in reading and 10.8 in composition. How many times greater is the former than the latter gain?

8. Four problems, 1, 2, 3, and 4, are solved by 50%, 60%, 70%, and 80%, respectively, of a large group. Compare the difference in difficulty between 1 and 2 with the difference in difficulty between 3 and 4.

9. In a college the 10 grades $A+$, A, $A-$; $B+$, B, $B-$; $C+$, C, $C-$; and D are given. On the assumption that ability in mathematics is distributed normally, how many men in a group of 500 Freshmen should receive each grade?

10. Five problems are passed by 15%, 34%, 50%, 62%, and 80% of a large unselected group. If the zero point of ability is taken at -3σ, what is the σ value of each problem as measured from this point?

11. In a large group of competent judges, 88% rank composition *A* as better than composition *B*; 65% rank *B* as better than *C*. If *C* is known to have the *PE* value of 3.5 as measured from the zero composition, i.e., the composition of zero merit, what are the *PE* values of *B* and *A* as measured from this "zero"?

12. Twenty-five men on a football squad are ranked in order of merit from 1 to 25 for general playing ability by the coach. Assuming "normality" in the trait "general playing ability" transmute these ranks into units of amount on a scale of 100 points.

Answers

4. (*a*) 57.04%. (*b*) 749 in 10,000. (*c*) 52 in 10,000.
5. (*a*) 4.8%. (*b*) 25.76 and 33.72. (*c*) 21.95 and the lower limit of the distribution.
6. 30.65%.
7. 2.96 (approximately 3) times as great.
8. Difference between 1 and 2, $.25\sigma$; between 3 and 4, $.315\sigma$.
9.

Grades:	*A*+	*A*	*A*−	*B*+	*B*	*B*−	*C*+	*C*	*C*−	*D*
No. men receiving:	3	14	40	80	113	113	80	40	14	3

10. In order: 4.04; 3.41; 3.00; 2.69; 2.16.
11. *B*, 4.07*PE*; *A*, 5.82*PE*.
12.

Rank	Score	Rank	Score
1	89	13	50
2	80	14	48
3	75	15	46
4	71	16	44
5	68	17	42
6	65	18	39
7	63	19	37
8	61	20	35
9	58	21	32
10	56	22	29
11	54	23	25
12	52	24	20
		25	11

CHAPTER III

THE RELIABILITY OF MEASURES

I. What is Meant by the Reliability of a Measure

By the "true" measure of an individual's capacity in any trait, as for example, the true measure of his height, reaction, time, or intelligence, we mean the average of an *infinite number* of measurements of the given capacity made under precisely the same conditions. Obviously, in actual practice, we can never deal with true measures as thus defined—for usually we must be satisfied with a single measure, or at best with a comparatively few measures of the given trait. We can, however, measure the amount by which an obtained measure "most probably" varies from its corresponding true measure; and this measure of "probable divergence" serves as an index of the *reliability* of the obtained measure—of how good an approximation it is of the true measure.

In like manner, the reliability of an obtained measure of a group is determined by finding the probable divergence of the obtained measure from the true measure of the group. The true measure of a group—as for example the true average or the true σ—is defined as that measure obtained by taking into account *all* of the members of the group, and the true measure of difference between two groups is the difference between their true means or medians. To show just what is meant by the "true measure" of a group, let us suppose that we could measure the height of every 12 year old boy in the United States. If from this frequency distribution of heights, we should calculate a measure of central tendency and a measure of variability—the average and σ for example—this average would be the true average height of 12 year old

boys in the United States, and the σ would be the true measure of scatter around this average. In the same way, if we could measure the height of every 12 year old girl in the United States, it would be possible to secure the true average height, and the true variability around it, of 12 year old girls in this country. Moreover, knowing the true average height of 12 year old boys and the true average height of 12 year old girls, it would be a very simple matter to find the true difference between the average height of 12 year old boys and 12 year old girls in the United States.

Unfortunately it is rarely, if ever, possible to measure *all* of the individuals in a group or "population," and it is, of course, impossible to take an infinite number of measures of a given individual. We must be content, therefore, to deal with "samples" selected from the total number of possible measures; and, as a result, due to slight differences in the samples chosen, measures of central tendency and variability are often larger or smaller than their corresponding true measures. Hence, whenever we have measured an individual or a group, we must ask ourselves this question: "How reliable a measure of capacity have I secured? How well does it 'represent' the true measure which I should get from a very large (infinite) number of measures of this individual—or from measuring *all* of the individuals in the population from which my group is taken?" This question will often lead to a second: "How many measurements must I make in order to get a result which shall meet a certain standard of reliability, i.e., show a probable divergence from the true result which is less than some given amount?"

The purpose of the following sections is to develop methods which will enable us to answer these questions. First, the reliability of the mean and median will be considered; then the reliability of the measures of variability; and finally the reliability of the difference between two measures.[1]

[1] The method of finding the reliability of a coefficient of correlation is given later on page 170.

II. The Reliability of Measures of Central Tendency

1. The Reliability of the Average or Mean

A. The Reliability of the Mean in Terms of its Standard Error ($\sigma_{av.}$)

Perhaps the simplest approach to the study of the reliability of the average is to examine the factors upon which the reliability of this measure must depend. Suppose that we wish to find the average score of college freshmen in the United States on Army Alpha. To measure the achievement of college freshmen in general, would require in strict logic that we test *all* of the freshmen in the United States. However, this is a well-nigh impossible task, and hence we must be satisfied with taking the records of as *large* and *random* a sample of freshmen as we can secure. This means that we cannot use freshmen from only a single institution or from only one section of the country, and that we must guard against selecting only those with low or high scholastic records. The more successful we are in getting an " unselected " group the more nearly representative will this group be of *all* of the freshmen in the country. Evidently, therefore, the reliability (the " representativeness ") of an average depends, for one thing, on how impartially we have selected our sample.

Granted a fair sample, the reliability of an average can be shown to depend upon two characteristics of the distribution, (1) the number of cases, and (2) the variability or spread of the measures within the sample.

(1) It is clear that the number of cases must influence the stability of an average, since the addition of even one extra measure to a series will bring about a change in the average unless the additional case happens to coincide with it exactly. Moreover, the addition of one case to a set of 10 measures will cause a greater change in the obtained average—written " $\text{average}_{(obt.)}$"—than the addition of one extra case to a set of 1000 measures, as each case counts for less in the larger

group. It has been shown empirically, as well as theoretically,[1] that the reliability of an $\text{average}_{(\text{obt.})}$ will increase, not in proportion to the number of measures upon which it is based, but rather in proportion to the *square root* of the number of measures. Thus the $\text{average}_{(\text{obt.})}$ of 25 measures of a variable quantity is not 25 times, but $\sqrt{25}$ or 5 times as reliable as a single measure of the quantity. And in like manner, the average of 36 cases is not 4 times as reliable as the average of 9 cases, but only twice as reliable—since $\sqrt{36}$ divided by $\sqrt{9}$ equals 2.

(2) In addition to the *size* of the sample, the reliability of an average must depend also upon the variability of the separate measures around the obtained average. If the σ of the distribution is large, the separate measures tend to scatter widely from the average, and we are unable to say where those cases in the population which we have not measured will most probably fall: whether they will be close to, or far from the obtained average. On the other hand, if the σ is small we may be fairly certain that unmeasured cases will fall fairly close around the average. For this reason, the reliability of an obtained average depends upon the size of its σ—and as σ increases, the reliability decreases.

We find, then, that the reliability of an average depends *first* upon our having selected a fairly representative sample from the larger group—or population—which we are studying. When this condition has been met, and only then, the reliability of an average can be measured mathematically in terms of its standard error—in terms of the number of cases, and the σ of the distribution (written $\sigma_{(\text{dis.})}$). The formula for the standard error of an average or mean, written $\sigma_{\text{av.}}$ is

$$\sigma_{\text{av.}} = \frac{\sigma_{(\text{dis.})}}{\sqrt{N}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

[1] Yule: *An Introduction to the Theory of Statistics*, 1919, p. 257. For results of experiment, see Fullerton and Cattell: *On the Perception of Small Differences*, Publications of the University of Pennsylvania, Philosophical Series 2, 1892.

This is one of the most important—and most often used—of the reliability formulas. Note that a decrease in $\sigma_{(\text{dis.})}$, or an increase in the size of N will cause the standard error to become smaller numerically. A decrease in $\sigma_{\text{av.}}$ means that the probable divergence of the obtained average from the true is just so much less; hence the reliability of an average$_{(\text{obt.})}$ increases as $\sigma_{\text{av.}}$ decreases.

A problem will illustrate the value and use of formula (13).

> *Problem* (1)—In 1883, the Anthropometric Committee of the British Association found the average height of 8585 adult males in the British Isles to be 67.46 inches with a σ of 2.57 inches.[1] How reliable is this average? What is its probable divergence from the average which would have been secured had *all* adult males in the British Isles been measured?

Applying formula (13) the standard error of the mean, $\sigma_{\text{av.}}$, is found to be .0277 inch. This result is interpreted in the following way. The chances are 6826 in 10,000 or 68 in 100 that the obtained average of 67.46 inches does not *diverge* from the true average by more than $\pm 1\sigma_{\text{av.}}$, i.e., by more than $\pm .0277$ inch. Stated in another way, the chances are 68 in 100 that the true average lies within the limits $67.46 + .0277$ and $67.46 - .0277$, or between 67.488 and 67.432 inches. We can be practically certain that the true mean lies within the limits $67.46 \pm 3 \times .0277$ $(\pm 3\sigma_{\text{av.}})$, or between 67.543 and 67.377 inches (see Table X for σ values).

Just how the standard error measures the reliability of an average may be shown most clearly, perhaps, by an illustration. Suppose that we have measured the heights of 1000 groups of men, each group containing 8585, the groups or samples chosen at random from the general population. The 1000 averages obtained from these groups will tend to differ slightly from one another due to so-called errors of sampling (see page 143) and hence not all samples will represent with equal accuracy the population from which they have been

[1] Yule, *An Introduction to the Theory of Statistics*, 1919, pp. 112 and 141.

drawn. Now suppose, further, that it were possible to secure the average height of the *entire male population* of the British Isles. If we should subtract this true mean from each one of the 1000 obtained means, obviously we would get 1000 differences, and these 1000 "measures" (differences) would—according to the best assumption that we can make—follow the normal probability curve (see page 83). In this hypothetical distribution of differences, we should have relatively *few large* plus or minus deviations, and a relatively *large* number of *small* plus, *small* minus, and *zero* deviations—in short, the obtained means would hit close to the true mean more often than they would miss it.

The average of this distribution of differences would fall (most probably) at 0; for other things being equal, this will be the difference most often obtained—the maximum frequency—in subtracting the true from the obtained means. The σ of this distribution is given by the formula $\frac{\sigma_{(\text{dis.})}}{\sqrt{N}}$. In other words, the standard error of the mean measures the spread of the differences (obtained-true) around 0 as a central tendency; and for this reason $\sigma_{\text{av.}}$ is a measure of the probable divergence of the obtained average from its corresponding true average.

These results are represented graphically in Diagram XV, Fig. 1. The 1000 differences between the 1000 obtained means and the true mean are shown arranged into a normal frequency distribution with mean at 0, and σ equal to .0277. The heights of the different ordinates represent the frequency of the various obtained-true differences: the height of the maximum ordinate at the mean is the zero difference. Now we know that the σ of a normal distribution includes the middle 68.26% of the cases, when measured off in the plus and minus directions from the mean. Hence we may say that the chances are 68 in 100 that the difference between the obtained mean of 67.46 inches and the true mean will not be greater than $\pm$.0277 inch. Or, as stated above, there are 68 chances in 100 that the true average

lies within the limits $67.46 + .0277$ and $67.46 - .0277$, or between 67.488 and 67.432 inches. Furthermore, we can be practically sure that the true average will fall within the limits $\pm 3\sigma_{av.}$ from the mean. Three times $\pm .0277$ is $\pm .0831$; and accordingly there are 9973 chances in 10,000 (see Table X) that the true average lies within the limits $67.46 \pm .0831$, or between 67.543 and 67.377 inches.

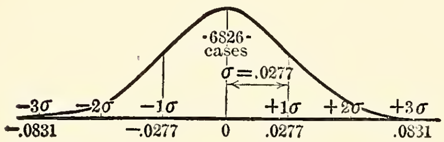


FIG. 1

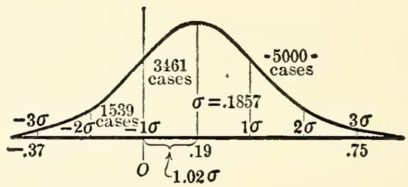


FIG. 2

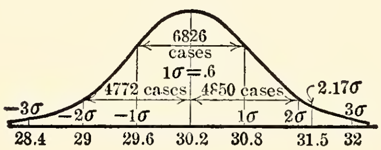


FIG. 3

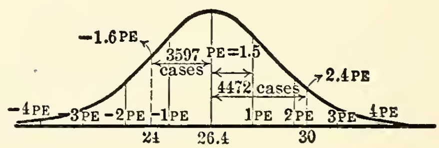


FIG. 4

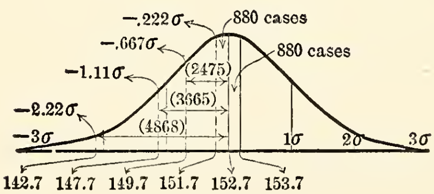


FIG. 5

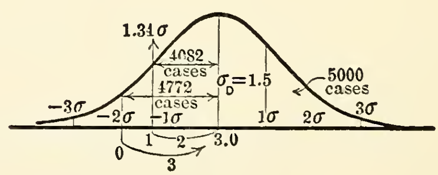


FIG. 6

DIAGRAM XV

The average height of our sample of 8585 British males has been found to be 67.46 inches with a standard error of .0277 inch. Let us now proceed to the second question stated on page 119, viz., "How many measurements must I make in order to get a result whose probable divergence from the true result is less than some given amount?" Suppose, for example, that we wish to secure an average which is *twice* as reliable as the average we now have—how many cases will be required? Assuming that the spread in the increased group,

i.e., $\sigma_{(\text{dis.})}$, remains approximately the same, all that we need do in order to cut the standard error in two and thus double the reliability, is to place a 2 in the denominator of the fraction $\frac{2.57}{\sqrt{8585}}$. But $2\sqrt{8585}$ becomes $\sqrt{4 \times 8585}$ when the 2 is placed under the radical, and, accordingly, it is evident that 8585 must be multiplied by 4 in order to make $\sigma_{\text{av.}}$ just 1/2 its original size. By analogy, to double the reliability of any average we must multiply N by 4; to triple the reliability, by 9, etc. Assuming substantially the same $\sigma_{(\text{dis.})}$, the average obtained from 400 cases is twice as reliable as the average got from 100, and the average from 900 cases three times as reliable as that from 100 cases.

B. The Reliability of the Mean in Terms of the *PE* of the Average

In measuring the reliability of an average the *PE* of the average—written $PE_{(\text{av.})}$—may be used instead of the $\sigma_{\text{av.}}$ The $PE_{(\text{av.})}$ is interpreted in exactly the same way as the $\sigma_{(\text{av.})}$. Its formula is derived simply by multiplying formula (13) by .6745 (see page 121):

$$PE_{(\text{av.})} = \frac{.6745\sigma_{(\text{dis.})}}{\sqrt{N}} \quad . \quad . \quad . \quad . \quad . \quad . \qquad (14)$$

Applying this formula to our problem of heights $PE_{(\text{av.})}$ is found to be .0187 inch. The chances are even, therefore, that the obtained average of 67.46 inches does not differ from the true average by more than $\pm .0187$ inch. Moreover, since $\pm 4PE$ includes practically all of the cases in a normal distribution, we may be certain (the chances are 99 in 100) that the true average lies within the limits $67.46 \pm 4 \times .0187$, or between 67.39 and 67.53 inches (see Table XI for *PE* values).

A comparison of the extreme limits within which we may be practically sure that the true average will lie shows that the values of these limits differ slightly when $\pm 4PE$ instead of $\pm 3\sigma$ are taken as limiting points [see Problem (1) above].

This discrepancy is due to the fact that $\pm 3\sigma$ takes in 9973 of the 10,000 cases in the normal distribution, while $\pm 4PE$ takes in but 9930 cases (see Tables X and XI). The σ limits, therefore, contain 43 more cases than the PE limits, and while 43 cases in 10,000 may seem to be an insignificant number—and is insignificant if taken from the middle of the distribution—even so few cases as this have considerable importance at the extremes of the distribution. This may be seen in the fact that we must take $\pm 4.45PE$, in order to have our PE limits correspond exactly to $\pm 3\sigma$, since these limits include 9974 cases in 10,000.

It is customary, however, in measuring reliability to use $\pm 4PE$ instead of $\pm 4.45PE$ as limits of practical certainty. In the first place, $\pm 4PE$ mark off limits within which the chances are very great—9930 in 10,000—that the true average will fall. And furthermore, the slight increase in reliability got by using $\pm 4.45PE$ instead of $\pm 4PE$ is not usually sufficient to offset the greater convenience of the latter figure.

2. The Reliability of the Median

The formulas for measuring the reliability of an obtained median are easily derived from those for measuring the reliability of the mean. The $\sigma_{(\text{mdn.})}$ and $PE_{(\text{mdn.})}$ are 1.25331, or roughly 5/4, times the $\sigma_{\text{av.}}$ and $PE_{(\text{av.})}$ respectively.

$$\sigma_{(\text{mdn.})} = \frac{5}{4} \cdot \frac{\sigma_{(\text{dis.})}}{\sqrt{N}}, \qquad (15)$$

$$PE_{(\text{mdn.})} = \frac{5}{4} \cdot \frac{.6745 \times \sigma_{(\text{dis.})}}{\sqrt{N}} = \frac{.8454\sigma_{(\text{dis.})}}{\sqrt{N}}, \qquad (16)$$

or

$$PE_{(\text{mdn.})} = \frac{5}{4} \cdot \frac{Q}{\sqrt{N}}.^{1} \qquad (16a)$$

Formulas (15), (16), and 16a) are all used and interpreted in the same way as the reliability formulas for the average or

[1] This formula should be used when Q and not σ is given.

mean. A problem will serve to show how the reliability of the median is found.

> *Problem* (2)—Measurement of 801 12 year old boys on the Trabue Language Scale A[1] gave the following results: Median $=21.4$; $Q=4.9$. What is the reliability of this median? How close is it to the true median score of 12 year old boys?

From formula (16*a*) the $PE_{(\text{mdn.})}$ is found to be .2164. The chances are 50 in 100, therefore, that the true median does not differ from 21.4 by more than $\pm .2164$. We may be practically certain that the true median lies within the limits $21.4 \pm 4 \times .2164$, or between 22.27 and 20.53.

Since $\sigma_{(\text{mdn.})}$ and $PE_{(\text{mdn.})}$ are both larger—approximately 1.25 times—than the corresponding measures of reliability of the average$_{(\text{obt.})}$, it is clear that the obtained average is always more reliable than the obtained median of the same group. For this reason the average is used whenever the highest reliability is sought (see page 50).

III. The Reliability of Measures of Variability

1. The Standard Deviation, or σ

We have seen that the reliability of an obtained average or obtained median is found by determining the probable divergence of the obtained from the true measure. In the same way, the reliability of an obtained σ or an obtained Q is measured by the probable divergence of this measure from the true σ or the true Q, viz., the σ or the Q which we should get from *all* possible measures of the trait in question. The formula for finding the reliability of an obtained σ is

$$\sigma_\sigma = \frac{\sigma_{(\text{dis.})}}{\sqrt{2N}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

In Problem (1), page 122, we found that for 8585 adult British males, the obtained σ—the σ taken around the

[1] Trabue, M. R., *Completion Test Language Scales*, 1916, p. 15.

average$_{(obt.)}$ of 67.46 inches—was 2.57 inches. The question may well be asked: how reliable is this σ? How well does it represent the true σ which we should get if deviations could be taken from the true average? Substituting for $\sigma_{(dis.)}$ and N in formula (17), the value of σ_σ is found to be .0196 inch. This means that the chances are 68 in 100 that 2.57 inches does not differ from the true σ by more than $\pm$.0196 inch; and that the chances are 997 in 1000 that the $\sigma_{(dis.)}$ does not differ from the true σ by more than $3\times\pm.0196$ or $\pm$.0588 inch. We can be practically certain, then, that the true σ lies within the limits $2.57\pm.0588$, or between 2.63 and 2.51 inches.

2. The Quartile Deviation, or Q

The reliability of the Q of a distribution is found from the formula,

$$\sigma_Q=\frac{1.11\times\sigma_{(dis.)}}{\sqrt{2N}}, \quad \ldots\ldots\ldots \quad (18)$$

or in terms of Q,

$$\sigma_Q=\frac{1.65\times Q}{\sqrt{2N}}. \quad \ldots\ldots\ldots \quad (18a)$$

The 801 12 year old boys who took the Trabue Completion Test, Scale A (see page 127), had a median score of 21.4 points with a Q of 4.9 points. What is the reliability of this Q? From formula (18a) σ_Q is found to be .202. The chances are 68 in 100, therefore, that 4.9, the obtained Q, does not differ from the true Q by more than $\pm$.202 point. And the chances are 9973 in 10,000 that the true Q lies within the limits $4.9\pm 3\times.202$, or between 5.5 and 4.3 points.

IV. The Reliability of the Difference between Two Measures

1. The Reliability of the Difference between Two Averages

A. The Reliability of the Difference in Terms of the $\sigma_{(diff.)}$

Suppose that we wish to find whether there is any difference in the performance of 10 year old boys and 10 year old girls

on a certain general intelligence test. The usual method of attacking this problem is to select as large and as random a sample of 10 year old boys and 10 year old girls as possible; give them our test, compute the average scores, and find the difference between the two averages. If this difference is, let us say, several points in favor of the girls, such a result would be evidence (on the face of it) for believing that the average girl is better than the average boy. Before drawing this conclusion definitely, however, we should know how reliable the obtained difference is: what its probable divergence is from the true difference which we should get if we could subtract the true average of the boys from the true average of the girls.[1] Otherwise, if we compared the averages of other groups of boys and girls similarly selected as our groups, we might wipe out or even reverse the difference found. One formula for calculating the reliability of an obtained difference is

$$\sigma_{(\mathrm{diff.})} = \sqrt{\sigma^2_{(\mathrm{av.}\,1)} + \sigma^2_{(\mathrm{av.}\,2)}}, \quad . \quad . \quad . \quad . \quad . \qquad (19)$$

in which $\sigma_{\mathrm{av.}\,1}$ is the standard error of the first obtained average, $\sigma_{\mathrm{av.}\,2}$ is the standard error of the second obtained average, and $\sigma_{(\mathrm{diff.})}$ is the standard error of the difference between the two averages. Thus to find the reliability of the difference between two averages, we must first know the reliability of the averages themselves.

Let us illustrate the use and value of formula (19) by means of a problem.

Problem (3)—In a study of the intelligence of foreign born white draft during the Great War, a sample of 308 native born Germans and a sample of 325 native born Danes were found to test as follows on the "combined scale:"[2]

Country of Birth	No. of Cases	Average Score	$\sigma_{(\mathrm{dis.})}$
Germany....................	308	13.88	2.43
Denmark....................	325	13.69	2.23

[1] Simpler methods of studying the significance of the difference between two averages are given in Chapter I, p. 40.

[2] The combined scale was made up of the 8 Alpha tests, the Stanford-Binet, and tests 4, 5, 6, and 7 of Beta. The maximum score was 25.

The difference between the two obtained averages is seen to be .19 in favor of the Germans. Is this a reliable difference? Would further testing of other groups of Germans and Danes give approximately the same difference; or is it probable that the difference would be reduced to zero, or even reversed in favor of the Danes? Stated more exactly, what is the probable divergence of this difference from the true difference between Germans and Danes? To answer these questions, we must find the reliability of the averages of the Germans and the Danes, and from these the reliability of the difference between the averages.

By formula (13) the standard errors of the two averages are,

For Germans:

$$\sigma_{\text{av.}} = \frac{2.43}{\sqrt{308}} \quad \text{or} \quad .1385.$$

For Danes:

$$\sigma_{\text{av.}} = \frac{2.23}{\sqrt{325}} \quad \text{or} \quad .1237.$$

Substituting these values in formula (19) we have that

$$\sigma_{(\text{diff.})} = \sqrt{(.1385)^2 + (.1237)^2} = .1857.$$

The actual difference between the two averages is .19, therefore, and the standard error of this difference, $\sigma_{\text{diff.}}$ is .1857.

An obtained difference is interpreted in terms of its standard error in exactly the same way in which an obtained average is interpreted in terms of its standard error. Thus we may say that the chances are 68 in 100 that the obtained difference of .19 does not diverge from the true difference by more than $\pm .1857$; and that the chances are 99 in 100 that .19 does not differ from the true difference by more than $3 \times \pm .1857$—by more than $\pm .56$ (see Table X).

To sum up our findings so far, we may be almost certain that the true difference between the averages of the Germans and Danes lies within the limits $.19 \pm .56$ or between $-.37$ and $+.75$. Note that the lower limit of this range is negative,

and in consequence there is at least *some* chance that the true difference is less than zero—that the average of the Danes will sometimes actually be *higher* than that of the Germans. In spite of the obtained difference in favor of the Germans, we cannot be 100% sure that the true difference between the average German and the average Dane is greater than zero.

Just what then, it may be asked, *are* the chances of a true difference greater than zero between Germans and Danes? Before answering this question, let us digress for the moment to consider the following hypothetical situation.[1] Suppose that we could secure the averages of 1000 groups of native born Germans and 1000 groups of native born Danes on the combined scale, the samples selected at random from the general population of native born Germans and Danes and roughly of the same size as the samples we have. Suppose further, that these groups could be paired off so that we should have 1000 differences between the obtained averages of Germans and Danes, these hypothetical differences corresponding to the actually obtained differences of .19. Now according to the best assumption that we can make this distribution of differences would follow the normal probability curve; the lower limit of the distribution would be at $-.37$, the upper limit at .75 and the mean at .19 as shown in Diagram XV, Fig. 2. The mean is taken at .19 because this is the difference actually obtained, and hence may be fairly taken as the most probable. Again, the chances are even that any other obtained difference will be greater or less than .19; and accordingly, the logical place for this difference would seem to be at the mean. The σ of this distribution of differences is .1857, the $\sigma_{\text{diff.}}$.

Now to determine the chances that the true difference between Germans and Danes is greater than zero, we divide .19, which is the distance of the *mean difference* from the zero difference, by .1857, the σ of the difference-distribution. This tells

[1] The argument here which differs somewhat from that on page 123 is believed to be better adapted to the present illustration than the other. The two are essentially the same, however.

us how far the zero difference is *below* the mean in σ *terms*. $\frac{.19}{.1857}$ is 1.02σ, and from Table X we find that in the normal curve 3461 cases in 10,000 lie between the mean and 1.02σ. Adding in the 5000 cases above the mean (see Digram XV, Fig. 2) and translating cases over into "chances," it is clear that the chances are 8461 in 10,000 that the true difference between the averages of Germans and Danes is greater than zero. We may be practically certain, therefore, when we compare groups of Germans and Danes on the combined scale, that 84 times in 100 or 4 times in 5, the difference between the average scores will be in favor of the Germans. This answers the question put on page 130: "What are the chances of a true difference greater than zero between the Germans and Danes?"

The obtained difference of .19 is sufficiently large to insure considerably more than an even chance of a true difference between Germans and Danes. It is not large enough, however, to guarantee that the Germans will *always* score higher, on the average, than the Danes. The further question arises, therefore:—how much difference would be required to insure *absolute* reliability,—to guarantee that the Germans will *always* lead the Danes. This question is easily answered with the help of Fig. 2. If the point -3σ below the mean (the point taken at $-.37$) were the zero-difference point, we should then be practically certain, since the whole curve of differences would lie to the *right* of this point, of a true difference *always* greater than zero. To accomplish this, however, i.e., to shift the zero-difference point down to $-.37$, the mean difference would have to be $.37+.19$ or $.56$. This new difference (D) divided by $\sigma_{\text{diff.}}$ would equal $\frac{.56}{.1857}$ or 3σ, and the chances would then be 9986.5 in 10,000 that the true difference between Germans and Danes on the combined scale will *always* be greater than zero.

We may summarize the preceding paragraphs as follows. The obtained difference between the averages of the Germans

and Danes on the combined scale is found to be .19, or 1/3 (approximately) of what it should be, (.56) to insure a completely reliable difference. The obtained difference is large enough, however, to guarantee that 4 times in 5 the average score of the native born Germans will be higher than the average score of the native born Danes.[1]

Once we understand what the $\sigma_{\text{diff.}}$ formula means, the reliability of an obtained difference in terms of "chances that the obtained difference represents a true difference greater than zero" may be conveniently read from Table XIV. For example, when $D = .19$ and $\sigma_{\text{diff.}} = .1857$, so that $\frac{D}{\sigma_{\text{diff.}}} = 1.02$, we find at once from the table that the chances are 84 in 100 that the true difference is greater than zero. Moreover, since a $\frac{D}{\sigma_{\text{diff.}}}$ of 3 means practically complete reliability, we know that a $\frac{D}{\sigma_{\text{diff.}}}$ of 1.02 is $\frac{1.02}{3}$ or about 34% of what it should be in order to insure a difference always greater than zero.

It is usually customary to take a $\frac{D}{\sigma_{\text{diff.}}}$ of 3 as indicative of complete reliability, since -3σ includes practically all of the cases in the "distribution of differences" below the mean (see Diagram XV, Fig. 2). A $\frac{D}{\sigma_{\text{diff.}}}$ greater than 3 is to be taken as indicating just so much added reliability.

B. The Reliability of the Difference in Terms of the $PE_{(\text{diff.})}$

The reliability of the difference between two obtained means may be measured by the $PE_{(\text{diff.})}$ as well as by the $\sigma_{(\text{diff.})}$. The formula for $PE_{(\text{diff.})}$ is

$$PE_{(\text{diff.})} = \sqrt{PE^2_{(\text{av. 1})} + PE^2_{(\text{av. 2})}}, \quad . \quad . \quad . \quad (20)$$

in which $PE_{(\text{av. 1})}$ and $PE_{(\text{av. 2})}$ are the PE's of the two given ob-

[1] Assuming that the samples used represent adequately—at least as adequately as the present samples—the population of native born Germans and Danes.

TABLE XIV

To Find the Chances of a True Difference Greater than Zero, Given the Actual Difference between the Two Obtained Measures, and the $\sigma_{diff.}$

For example: a $\frac{D}{\sigma_{diff.}} = 1.3$ means that the chances are 90 in 100 that the true difference (the difference between the true measures) is greater than zero.

Note.—The "chances in 100" increase so slowly after 1.50 that the $\frac{D}{\sigma_{diff.}}$ column increases thereafter by .10 instead of by .05.

$\frac{D}{\sigma_{diff.}}$	Chances in 100	$\frac{D}{\sigma_{diff.}}$	Chances in 100
.00	50	1.15	87
.05	52	1.20	88
.10	54	1.25	89
.15	56	1.30	90
.20	58	1.35	91
.25	60	1.40	92
.30	62	1.45	93
.35	64	1.50	93
.40	65	1.60	94
.45	67	1.70	96
.50	69	1.80	96
.55	71	1.90	97
.60	73	2.00	98
.65	74	2.10	98
.70	76	2.20	99(98.6)
.75	77	2.30	99(98.9)
.80	79	2.40	99(99.2)
.85	80	2.50	99(99.4)
.90	82	2.60	99(99.5)
.95	83	2.70	100(99.7)
1.00	84	2.80	100(99.74)
1.05	85	2.90	100(99.8)
1.10	86	3.00	100(99.9)

tained averages. Formula (20) is interpreted in exactly the same manner as formula (19)—a problem will illustrate its use.

Problem (4)—On the two halves of the Woodworth-Wells Substitution Test[1] timed separately, 200 Barnard Freshmen made the following records:

	Average (Secs.)	σ(dis.)
First half.............	65.51	11.13
Second half...........	60.32	12.04

[1] Carothers, F. E., *Psychological Examination of College Students*, Archives of Psychology, 46, 1921, p. 36.

TABLE XV

To Find the Chances of a True Difference Greater than Zero, Given the Actual Difference between the Two Measures and the $PE_{\text{diff.}}$

For example: a $\frac{D}{PE_{\text{diff.}}} = 1.10$ means that there are 77 chances in 100 that the true difference (the difference between the true measures) is greater than zero.

Note.—The "chances in 100" increase so slowly after 2.0 that the $\frac{D}{PE_{\text{diff.}}}$ column increases thereafter by .10 instead of .05.

$\frac{D}{PE_{\text{diff.}}}$	Chances in 100	$\frac{D}{PE_{\text{diff.}}}$	Chances in 100
.00	50	1.55	85
.05	51	1.60	86
.10	53	1.65	87
.15	54	1.70	87
.20	55	1.75	88
.25	57	1.80	89
.30	58	1.85	89
.35	59	1.90	90
.40	61	1.95	91
.45	62	2.00	91
.50	63	2.10	92
.55	64	2.20	93
.60	66	2.30	94
.65	67	2.40	95
.70	68	2.50	95
.75	69	2.60	96
.80	71	2.70	97(96.6)
.85	72	2.80	97
.90	73	2.90	97(97.5)
.95	74	3.00	98(97.9)
1.00	75	3.10	98
1.05	76	3.20	98(98.5)
1.10	77	3.30	99(98.7)
1.15	78	3.40	99(98.9)
1.20	79	3.50	99
1.25	80	3.60	99
1.30	81	3.70	99
1.35	82	3.80	99(99.5)
1.40	83	3.90	100(99.6)
1.45	84	4.00	100(99.7)
1.50	84		

Is this gain in time from the first to the second half of the test sufficiently large to indicate a true difference in the time required to learn the key after practice, or would further testing with other groups probably reduce, or even reverse, the gain?

First, to find the probable errors of the two averages:

First half:

$$PE_{(\text{av. }1)} = \frac{.6745 \times 11.13}{\sqrt{200}} = .5310. \quad \text{By formula (14)}$$

Second half:

$$PE_{(\text{av. }2)} = \frac{.6745 \times 12.04}{\sqrt{200}} = .5743. \quad \text{By formula (14)}$$

Substituting $PE_{(\text{av. }1)}$ and $PE_{(\text{av. }2)}$ in formula (20) we have

$$PE_{(\text{diff.})} = \sqrt{(.5310)^2 + (.5743)^2} = .7822.$$

The obtained difference, D, is 5.19 and the $PE_{(\text{diff.})}$ is .7822. Therefore, $\frac{D}{PE_{(\text{diff.})}}$ is 6.64, and since we find from Table XV (to be read exactly like Table XIV) that a $\frac{D}{PE_{(\text{diff.})}}$ of 4 indicates complete reliability, it follows that our obtained difference is not only completely reliable, but is $2.64PE(6.64-4.00)$ or about 66% larger than it need be in order to insure a true difference greater than zero.

Just as it is customary to take a $\frac{D}{\sigma_{\text{diff.}}}$ of 3 as indicative of complete reliability, so a $\frac{D}{PE_{(\text{diff.})}}$ must be at least 4 in order to insure complete reliability.

2. The Reliability of the Difference between Two Medians

The two formulas (19) and (20), used in finding the reliability of the difference between two means, may be used also for finding the reliability of the difference between two medians when written:

$$\sigma_{(\text{diff.})} = \sqrt{\sigma^2_{(\text{mdn. }1)} + \sigma^2_{(\text{mdn. }2)}}, \quad . \quad . \quad . \quad . \quad . \quad (21)$$

and

$$PE_{(\text{diff.})} = \sqrt{PE^2_{(\text{mdn. }1)} + PE^2_{(\text{mdn. }2)}}, \quad . \quad . \quad . \quad . \quad (22)$$

We may illustrate these formulas by a problem:

Problem (5)—The following results were obtained from a group of 12 year old boys and a group of 12 year old girls—Grades III to VIII inclusive—on the Trabue Language Scale A.[1]

	N	Median	Q
Boys...........	801	21 40	4.9
Girls...........	448	22.80	5.3

The actual difference between the two medians is 1.4 points in favor of the girls. Assuming that the two groups are fairly unselected, is this difference sufficiently large to insure a true difference greater than zero in favor of the girls?

Since the measure of variability given is the Q, we shall use the formula for $PE_{(\text{diff.})}$. First, to find the reliability of the two medians:

For girls: $PE_{(\text{mdn.})}=\frac{5}{4}\cdot\frac{5.3}{\sqrt{448}}=.3130.$ By formula (16*a*)

For boys: $PE_{(\text{mdn.})}=\frac{5}{4}\cdot\frac{4.9}{\sqrt{801}}=.2164.$ By formula (16*a*)

Substituting in (22) we have,

$$PE_{(\text{diff.})}=\sqrt{(.3130)^2+(.2164)^2}=.3805$$

The obtained difference is 1.4 and the $PE_{(\text{diff.})}$ is .3805. Therefore, $\frac{D}{PE_{(\text{diff.})}}$ is 3.68, and from Table XV we find that the chances are 99.3 in 100 that there is a difference greater than zero between the true median scores of 12 year old boys and girls. The obtained difference is only 92% $\left(\frac{3.68}{4.00}\right)$ of what it should be conventionally in order to guarantee complete reliability. However, it is sufficiently high to be taken—for all practical purposes—as completely reliable.

[1] *Completion-Test Language Scales*, 1916, p. 15.

V. Some Problems Which Involve Measures of Reliability

This Section is designed to illustrate a variety of problems which require in their solution the reliability formulas given in this Chapter and the frequency tables. For quick reference later, each group of examples is preceded by a general statement of the essential problem involved.

A. To Find the Probability That the True Average is Greater or Less than Some Designated Point on the Scale, or That it Falls within Given Limits

Problem (1)—Given $\text{Average}_{\text{obt.}} = 30.2$. $\sigma_{(\text{dis.})} = 6.00$. $N = 100$. On the assumption that this sample is fairly representative of the population from which it is drawn, (*a*) what is the reliability of the obtained average? (*b*) What are the chances that the true average is less than 29? (*c*) greater than 31.5? (*d*) that the true average lies between 28 and 31?

(*a*) From formula (13) we find that the $\sigma_{\text{av.}}$ is .6; hence the chances are 68 in 100 that the obtained average does not diverge from the true average by more than $\pm .6$, and that the true average falls between the limits 30.8 and 29.6. Moreover, the chances are 99.7 in 100 that 30.2 does not diverge from the true average by more than $\pm .6 \times 3$ or ± 1.8; i.e., that the true average falls within the limits 28.4 and 32.

These results are represented graphically in Diagram XV, Fig. 3. This normal probability distribution represents the distribution of means that we should expect to get from a large number of random samples, selected in the same way as the sample we have.[1] The central tendency of this hypothetical distribution of means is taken at 30.2, the actually obtained, and hence the most probable, mean. The standard deviation of the distribution is .6, the standard error of the given obtained mean.

(*b*) What are the chances that the true mean is less than 29?

[1] See the discussion on pages 122–123.

29 lies 1.2 points or 2σ below the obtained mean of 30.2 (see Fig. 3). From Table X, we find that 4772 cases in 10,000 fall between the mean and 2σ in a normal distribution; and, accordingly, $5000-4772$ or 228 cases must lie *below* 2σ. The chances are 228 in 10,000, therefore, that the true mean lies below—is less than—29.

(*c*) What are the chances that the true mean is greater than 31.5? This score is 1.3 points or 2.17σ above the obtained mean. There are 4850 cases in 10,000 between the mean and 2.17σ in a normal distribution: and $5000-4850$ or 150 cases above this point. Hence the chances are 150 in 10,000 or about 2 in 100 that the true mean is greater than 31.5 (i.e., lies above 2.17σ).

(*d*) What are the chances that the true mean lies between 28 and 31? 28 is 2.2 points or -3.67σ from the mean; and 31 is .8 of a point or 1.34σ from the mean. Between the mean and -3.67σ in a normal distribution are 4999 cases in 10,000, and between the mean and 1.34σ are 4099 cases in 10,000. Within the interval from -3.67σ to 1.34σ, therefore, we find $4999+4099$ or 9098 cases. Stated as chances, there are about 91 chances in 100 that the true average lies between 28 and 31.

> *Problem* (2)—Given $\text{Average}_{(\text{obt.})}=26.4$. $PE_{(\text{av.})}=1.5$. What are the chances that the true average of the group of which the given group is a random sample is (*a*) as large as 30? (*b*) as small as 24?

As in Problem (1), this situation may be represented by a normal probability curve, with the mean at 26.4 and *PE* equal to 1.5 (see Diagram XV, Fig. 4).

(*a*) What are the chances that the true average of the group is as large as 30? 30 is 3.6 points or 2.4 *PE above* the obtained average of 26.4. There are 4472 cases in 10,000 between the mean and 2.4 *PE* in a normal distribution (Table XI); and $5000-4472$ or 528 cases above 2.4 *PE*, i.e., above 30. Hence the chances are 528 in 10,000 or about 5 in 100 that the true average is as large (or larger than) 30.

(*b*) What are the chances that the true average is small as 24? 24 lies 2.4 points or -1.6 *PE* from the mean. There are 3597 cases in 10,000 between the mean and -1.6 *PE* in a normal distribution, and $5000-3597$ or 1403 cases below -1.6 *PE*. The chances are 1403 in 10,000, therefore, that the true average is as small (or smaller than) 24.

B. To Find the Probability That the Divergence of an Obtained Measure from its True Measure Will be within Given Limits

Problem (3)—Given $\text{Average}_{(\text{obt.})}=152.7$ and $\sigma_{(\text{av.})}=4.5$. Find the probability that the given obtained average will not diverge (or vary) from the true, by more than (*a*) 1 point, (*b*) 3 points, (*c*) 5 points, (*d*) 10 points.

(*a*) This is essentially the same problem, expressed in a slightly different way, as the problems under A. To find the probability that the obtained average differs from the true by as much $+1$ or -1, we must find the chances that the true mean lies within the limits 152.7 ± 1, i.e. between 151.7 and 153.7. (This is shown in Diagram XV, Fig. 5). A deviation of ±1 point is a deviation of $\pm\frac{1}{4.5}$ or $\pm.222\sigma$ from the obtained mean. From Table X we find that 880 cases in 10,000 in a normal distribution fall between the mean and $+.222\sigma$ or $-.222\sigma$. Accordingly, 880×2 or 1760 cases fall within the interval $+.222\sigma$ to $-.222\sigma$, and the chances are 1760 in 10,000 that the obtained mean will not diverge from the true mean by more than ±1 point.

(*b*) Three points are $\pm\frac{3}{4.5}$ or $\pm.667\sigma$ from the mean. There are 2475×2 or 4950 cases within the interval $.667\sigma$ measured off to the right and left of the mean. Hence there are 4950 chances in 10,000 that the obtained mean will not diverge from the true mean by more than ±3 points.

(*c*) Five points are $\pm\frac{5}{4.5}$ or $\pm1.11\sigma$ from the mean. Hence there are 3665×2 or 7330 chances in 10,000 that the obtained

average will not differ from the true average by more than ± 5 points.

(*d*) Ten points are $\pm\frac{10}{4.5}$ or $\pm 2.22\sigma$ from the mean; and accordingly there are 4868×2 or 9736 chances in 10,000 that the obtained mean will not diverge from the true mean by more than ± 10 points.

C. To Find the Probability That the True Difference between the Measures of Two Groups is Greater or Less than a Given Amount

> *Problem* (4)—The difference between two obtained means is 3. $\sigma_{(\text{diff.})} = 1.5$. (*a*) What are the chances that the true difference between the means of the two groups is greater than 0? (*b*) greater than 1? (*c*) greater than 3?

(*a*) Zero difference is $\frac{3}{1.5}$ or 2σ below the mean of differences, viz., 3 (see Diagram XV, Fig. 6). There are 4772 cases in 10,000 between the mean of a normal distribution and 2σ. Accordingly, there are $5000+4772$ or 9772 chances in 10,000 that the true difference is greater than zero. (Note that this result may be read off directly from Table XIV—that $\frac{D}{\sigma_{\text{diff.}}} = 2$.)

(*b*) One is $\frac{2}{1.5}$ or 1.33σ below the mean. There are 4082 cases in 10,000 in a normal distribution between the mean and 1.33σ. The chances, therefore, are $5000+4082$ or 9082 in 10,000 that the true difference is greater than 1.

(*c*) What are the chances that the true difference is greater than 3? The obtained difference of 3 has been placed at the mean of differences as the obtained, and hence the most probable difference. The chances are even, therefore, or 50–50 that the true difference is greater (or less) than 3. Note that $\frac{D}{\sigma_{(\text{diff.})}}$ is $\frac{0}{1.5}$ or 0. (Table XIV.)

VI. Limitations to Reliability Formulas, and Cautions to be Observed in Interpreting Them

The formulas which have been given in this chapter for calculating the standard errors of obtained measures of central tendency and variability make use of only two characteristics of the distribution from which the measure has been obtained, viz., the σ (distribution)—the spread of the measures—and N, the number of cases. It is obvious that so far as the formulas themselves are concerned there is nothing which would prevent our finding a standard error for a measure obtained from any group. Such a general and uncritical application of reliability formulas, however, will almost surely lead to erroneous conclusions, and for this reason it is necessary to indicate briefly some of the limitations to reliability formulas as well as some cautions to be observed in interpreting results secured from them.

(1) In the first place, in interpreting standard errors we always make the assumption that measures obtained from *successive samples* are distributed according to the normal probability curve. This assumption is *only* true, however, when the number of cases is large; it is *not* valid when the sample is small. Hence the significance of a measure of reliability is conditioned upon our having a sufficiently large number of cases. If N is less than 25, there is little sense or justification in using reliability measures. One simple and practical method of judging whether the sample is "sufficiently" large is to continue taking independent measures or adding cases drawn at random, until the addition of extra cases fails to produce an appreciable fluctuation in the average or median. When this point is reached the sample is probably large enough to be taken as fairly representative of the larger group from which it has been drawn. As a corollary it must be recognized, however, that mere numbers are not in themselves a guarantee of a representative sample.

(2) A more serious limitation to the measures of reliability

arises from the fact that standard and probable errors of obtained measures can be assumed to measure *only* those errors which result from fluctuations due to "random sampling." An illustration will make this term clear. On page 122 we found that the obtained average height of 8585 adult British males was 67.46 inches with a standard error of .0277 inch. This means that the chances are 997 in 1000 that the true average height of British males lies between 67.54 and 67.38 inches. Now by "true average height" we mean the average height of *all* British males, from whom our group of 8585 is an attempted random sampling. If our group were perfectly representative, its average would equal the true average exactly. Except by chance, however, neither this sample nor another similarly selected, and approximately of the same size, will represent the entire population perfectly; and furthermore, it is extremely unlikely that the averages calculated from successive samples will equal each other. Nevertheless, if the samples are actually random, and there are no large constant errors present, the calculated averages will tend to vary around the true average of the whole group within a comparatively small range. Variations like these, which arise from the fact that we must generally work with samples instead of the whole population, are called "errors of sampling."

The function of the standard and of the probable errors is to give a measure of this sampling error, i.e., of the *probable* amount of deviation to be expected in an obtained measure from the corresponding true measure, as a result of working with a single sample. In other words, the standard or probable error measures the error made in taking a sample as representative of the larger group or population. If the standard error of a given mean is small, it does not follow that the obtained mean is highly reliable, *necessarily;* a small standard error indicates merely that the reliability is high, in so far as fluctuations due to differences in sampling are concerned.

Reliability formulas give no measure of the effects of errors due to other causes than those which arise from sampling.

Errors which arise from the failure to get a random sample, for example, are neither detected nor measured by these formulas. To illustrate this point, the average Army Alpha score made by 500 college men between the ages of 18 and 25 will not be representative of the male population of this age-range. College men form a highly selected group, and in consequence, other samples of 500 drawn at random from the male population between the ages of 18 and 25 will return very different results from that of the college group. These differences in average score cannot be attributed to errors of sampling; and to take this group as representative of the general male population between the ages of 18 and 25, and to calculate the standard error of its average will lead to an entirely erroneous idea of the intelligence of the general population. (The given sample might, of course, serve very well as a group representative of the population of college men.)

Other variations not measured by the reliability formulas arise from errors due to practice, fatigue, coachability of tests, faulty technique in giving and scoring tests, and, in fact, errors due to a bias of any sort. Standard errors calculated for measures secured from samples which contain such errors will always be of doubtful value.

The careful study of successive samples, retests when practicable, care in controlling conditions, and the use of objective checks whenever possible, will eliminate many of these troublesome and prolific sources of error. Assuming that constant errors are small or practically negligible, one of the simplest tests of the adequacy—the "representativeness"—of a sample consists in taking several other groups of approximately the same size from the general population. If the measures calculated from these groups are of very nearly the same size, we may be reasonably assured that we have representative samples. If the similarity is not fairly close, we must continue adding cases until the successive samples are approximately similar. Oftentimes more information may be secured in regard to the reliability of our measures in this

way than could be obtained from a blanket use of reliability formulas.

(3) In concluding this discussion, we should add one word in regard to the use of formulas which measure the reliability of the difference between two obtained measures, namely, $\sigma_{(\text{diff.})}$ and $PE_{(\text{diff.})}$. These formulas make allowance only for variable errors in the original measures—for errors which arise in sampling. Constant errors in the original scores and errors of the sort mentioned above are not detected, nor their influence measured. Furthermore, these formulas always assume that the measures or scores in the two series which are compared are *uncorrelated* (see page 288). These limitations must be borne in mind when using or interpreting differences in terms of the " true " difference. . . .

VII.—Summary of Reliability Formulas

1. The Reliability of Measures of Central Tendency

(1) The Average or Mean

1. $$\sigma_{(\text{aver.})}=\frac{\sigma_{(\text{dis.})}}{\sqrt{N}}. \qquad (13)$$

2. $$PE_{(\text{aver.})}=\frac{.6745\sigma_{(\text{dis.})}}{\sqrt{N}}. \qquad (14)$$

(2) The Median

1. $$\sigma_{(\text{mdn.})}=\frac{5}{4}\frac{\sigma_{(\text{dis.})}}{\sqrt{N}}. \qquad (15)$$

2. $$PE_{(\text{mdn.})}=\frac{5}{4}\frac{.6745\sigma_{(\text{dis.})}}{\sqrt{N}}. \qquad (16)$$

3. $$PE_{(\text{mdn.})}=\frac{5}{4}\frac{Q}{\sqrt{N}}. \qquad (16a)$$

2. The Reliability of Measures of Variability

(1) The Standard Deviation

1. $$\sigma_{\sigma}=\frac{\sigma_{(\text{dis.})}}{\sqrt{2N}}. \qquad (17)$$

(2) The Quartile Deviation

$$1. \quad \sigma_{(Q)} = \frac{1.11\sigma_{(\text{dis.})}}{\sqrt{2N}} \quad \text{(18)}$$

$$2. \quad \sigma_{(Q)} = \frac{1.65Q}{\sqrt{2N}} \quad \text{(18a)}$$

3. The Reliability of the Difference between Two Measures

(1) The Average

$$1. \quad \sigma_{(\text{diff.})} = \sqrt{\sigma^2_{(\text{aver. 1})} + \sigma^2_{(\text{aver. 2})}} \quad \text{(19)}$$

$$2. \quad PE_{(\text{diff.})} = \sqrt{PE^2_{(\text{aver. 1})} + PE^2_{(\text{aver. 2})}} \quad \text{(20)}$$

(2) The Median

$$1. \quad \sigma_{(\text{diff.})} = \sqrt{\sigma^2_{(\text{mdn. 1})} + \sigma^2_{(\text{mdn. 2})}} \quad \text{(21)}$$

$$2. \quad PE_{(\text{diff.})} = \sqrt{PE^2_{(\text{mdn. 1})} + PE^2_{(\text{mdn. 2})}} \quad \text{(22)}$$

PROBLEMS

Note: For uniformity in figuring "chances" in the following problems, take all σ and PE distances to three decimals and correct back to the second place. Count all fractions over one half as wholes and drop all under one half. For example, write 1.876σ as 1.88σ; .023 PE as .02 PE, etc.

1. Given that the obtained average is 26.4; σ is 3.2; N is 100.
   (*a*) What are the chances that the true average for the 10,000 from which the 100 cases measured are a random sampling will be greater than 27?
   (*b*) That it will be between 26 and 27?
   (*c*) What are the chances that the true variability will be between 3.1 and 3.3?
   (*d*) That the true variability will be less than 3.5?

2. Given: Median = 72.40. Q = 12.84. N = 81.
   (*a*) What are the chances that the true median of the population from which this random sample is drawn is above 75?
   (*b*) That it lies between 70 and 74?
   (*c*) What are the chances that the true Q is not greater than 15?
   (*d*) That it lies between 10 and 14?

3. Given: Av. $1 = 29.6$. $\sigma_{(\text{dis.})} = 3.54$. $N = 100$.
   Av. $2 = 28.4$. $\sigma_{(\text{dis.})} = 5.36$. $N = 225$.

   (*a*) Find the $\sigma_{\text{av.}}$ for both distributions.
   (*b*) Find the reliability of the difference between the means.
   (*c*) What difference would be completely reliable, assuming that the variability remains practically unchanged?

4. In Example 2, page 56, find the reliability of the difference between the means of distributions A and B [use the $\sigma_{(\text{diff.})}$].

5. $\text{Average}_{(\text{obt.})} = K$. $PE_{(\text{av.})} = 3.5$. What are the chances that the true average will not diverge from the obtained by more than (*a*) 1, (*b*) 3, (*c*) 10.

6. Given that Mdn. 1 − Mdn. $2 = 3.6$. $PE_{(\text{diff.})} = 3.0$.

   (*a*) What are the chances that true difference is less than 0?
   (*b*) That it is 1 or more?
   (*c*) What per cent is the obtained difference of the difference necessary for complete reliability?

7. Find the reliability of the average in

   (*a*) Example 4, page 116.
   (*b*) Example 5, page 116.

8. In a random sample of 100 cases each from the four groups A, B, C, and D, the following are obtained:

   A. Average = 101. $\sigma_{(\text{dis.})} = 10.0$.
   B. Average = 104. $\sigma_{(\text{dis.})} = 11.0$.
   C. Average = 93. $\sigma_{(\text{dis.})} = 9.6$.
   D. Average = 86. $\sigma_{(\text{dis.})} = 8.5$.

   What are the chances that, in general, the average of

   (*a*) the A's is better than the average of the B's.
   (*b*) the A's is 5 better than the average of the C's.
   (*c*) the A's is 10 better than the average of the D's.

   What are the chances that

   (*a*) a B will be better than the average A.
   (*b*) a B will be better than the average C.
   (*c*) a B will be better than the average D.

Answers

1. (*a*) 3 in 100.
   (*b*) 86 in 100.
   (*c*) 34 in 100.
   (*d*) 91 in 100.

2. (*a*) 16 in 100.
   (*b*) 55 in 100.
   (*c*) 90 in 100.
   (*d*) 71 in 100.

3. (*a*) $\sigma_{\text{av. }1} = .354$. $\sigma_{\text{av. }2} = .357$.
   (*b*) 99 chances in 100 of a true difference.
   (*c*) 1.51.

4. 92 chances in 100 of a true difference. (Table XIV).

5. (*a*) 15 in 100.
   (*b*) 44 in 100.
   (*c*) 95 in 100.

6. (*a*) 21 in 100.
   (*b*) 72 in 100.
   (*c*) 30%.

7. (*a*) $\sigma_{\text{av.}} = .0791$.
   (*b*) $PE_{\text{av.}} = .318$.

8. (*a*) 222 in 10,000.
   (*b*) 9846 in 10,000 or 99 in 100.
   (*c*) 9999.277 in 10,000 (100%).

   (*a*) 61 in 100.
   (*b*) 84 in 100.
   (*c*) 95 in 100.

CHAPTER IV

CORRELATION

I. What is Meant by Correlation

Up to this point in our discussion we have concerned ourselves chiefly with methods of computing statistical measures which shall represent in a reliable way the performance of an individual or a group in some defined capacity or trait. Frequently, however, it is of greater importance to examine the *relation* of some capacity, such as general intelligence, to some other capacity, such as musical ability, than to measure performance in a single trait alone. For example, we may ask whether there is any relation between general intelligence as measured by a standard intelligence test and scholastic achievement as measured by "grades" or "marks." Or, more specifically, we may inquire whether an individual who gives evidence of high general intelligence tends to outstrip the average individual in school work. Again, knowing the ability of an individual in one test, can we say anything about his ability in another and different test? Are certain abilities highly related, and others relatively independent? These questions, and others of the same general nature, are studied by the Method of Correlation.

The statistical device whereby relationship is expressed on a quantitative scale is called the "coefficient of correlation," and is designated by the letter "r."

Let us consider first the situation where the correlation is fixed and unchanging. We know that the circumference of a circle is always 3.1416 times its diameter, no matter how large or how small the circle, or in what part of the world we

find it. Each time that we increase or decrease the diameter of a circle, we increase or decrease the circumference by just 3.1416 times the same amount. In short, the relation is fixed and definite, and hence we say that the "correlation" between diameter and circumference is perfect, and that r is equal to 1.00. In like manner, if we find that 100 men take exactly the same arrangement in two tests, so that the man who ranks first (or highest) in the one ranks first in the other, the man who ranks second in the first test ranks second in the other, and that this one-to-one correspondence holds throughout the entire list, the correlation here is perfect also, for the relative position of each man is exactly the same in one test as in the other. The coefficient of correlation, r, is equal to 1.00.

Now let us consider the case where there is just no relation at all. Suppose that we have examined 100 college seniors on the Army Alpha test and on a tapping test. The average Alpha score for the whole group is 175, and the average tapping rate is 185 taps in 30 seconds. Suppose further, that when we divide our group into three equal parts, the average Alpha score of the upper one-third is 190, and the average tapping rate 184; the average Alpha score of the middle third is 175 with an average tapping rate of 186; and the average Alpha score of the lowest one-third is 160 with an average tapping rate of 185. Now clearly since the tapping rate is almost identical in all three groups, we should be unable to draw any conclusion from a man's tapping rate *alone* as regards his probable score on Alpha. An average tapping rate of, say, 185 to 190, is as liable to be found with an Alpha score of 150 as with one of 175 or even 200. We should be as well qualified, then, to estimate a man's Alpha score knowing only his tapping rate as we should be able to estimate it if all we knew about the man in question was that he had blue eyes and light hair. In either case our estimate would be no better than a guess. There is, therefore, little or no correspondence in the *degree* or *amount* of capacity possessed by a given individual in the traits measured by the two tests, and the

coefficient of correlation r will equal zero, which means that there is just no correlation present.

So far we have indicated that *perfect* relationship may be expressed by a coefficient of 1.00, and that just no relation by a coefficient of 0. Between these two limits we may have relations of varying degree, indicated by such coefficients as .30, .60, .90. In every case a coefficient between 0 and 1.00 implies some degree of *positive* association, the degree of association depending on size of the coefficient.

Relation may be *negative* as well as positive, however. That is, a large degree of one ability may be associated with a small degree of another, or vice versa. When this inverse relation is perfect, r equals -1.00. To illustrate, suppose that in a certain group of 25 boys, we find that the boy standing highest in Latin ranks lowest in Shop Work; that the boy who stands second in Latin stands next to the bottom in Shop Work; and that any given boy is found to stand exactly the same distance from the top of the group in Latin as he stands from the bottom of the group in Shop Work. Table XVI on p. 152 will illustrate the situation.

The correspondence here is fixed and definite enough, but the relation is *inverse*. Hence the correlation, while perfect, is *negative*, and the coefficient of correlation r equals -1.00. Negative coefficients may range all the way from -1.00 up to 0, just as positive coefficients range from 1.00 down to 0.

Coefficients of correlation, then, may range up and down on a scale which extends from -1.00 through 0 to $+1.00$. A positive correlation indicates a *positive* relation or correspondence; a zero correlation the *absence* of relation; and a negative correlation indicates an *inverse* relation. While for the sake of simplicity, we have illustrated above only perfect positive, perfect negative, and zero correlation, only rarely do we get coefficients at the extremes of the scale. In most cases calculated coefficients will be found at intermediate points, e.g., at .90, .20, $-.30$, etc. Such intermediate values as these are to be interpreted as " high " or " low " in a general way

depending upon how close they are to ± 1.00 or 0. A more complete discussion of the meaning of a correlation coefficient is given later on page 160.

TABLE XVI

To Illustrate a Correlation of -1.00

Boy	Standing in Latin	Standing in Shop Work
1	1	25
2	2	24
3	3	23
4	4	22
5	5	21
6	6	20
7	7	19
8	8	18
9	9	17
10	10	16
11	11	15
12	12	14
13	13	13
14	14	12
15	15	11
16	16	10
17	17	9
18	18	8
19	19	7
20	20	6
21	21	5
22	22	4
23	23	3
24	24	2
25	25	1

II. The Coefficient of Correlation:—What it is, and What It Does

1. The Coefficient of Correlation as a Ratio

Instead of taking up directly the method of computing an r, we shall first try in this section to give a clear notion of just what an r represents and how it measures relationship. The steps in the calculation of r by the "product-moment" method—the standard method—will then be given in detail in the next section.

Let us begin with Diagram XVI. This diagram, which is

DIAGRAM XVI

To Show How Correlation May be Expressed as a Ratio

Height in cms. (*Y*-variable)	Weight in Kgs. (*X*-variable) 45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	*Fy*	Av. wt.
189–185								1	1	82.5
184–180			1	3	3	4	2	3	16	71.3
179–175			4	11	6	3	2	2	28	66.4
174–170		2	9	11	8	2	1		33	62.8
169–165	1	5	7	10	3				26	59.2
164–160	1	2	7	1	2				13	57.9
159–155	1	1		1					3	54.2
Fx	3	10	28	37	22	9	5	6	120	
Av. ht.	162.5	166.5	169.8	172.8	173.6	178.6	178.5	181.7		

(A)

Weight	Av. ht. for given wt.
80–84	181.7
75–79	178.5
70–74	178.6
65–69	173.6
60–64	172.8
55–59	169.8
50–54	166.5
45–49	162.5

Range, $85-47.5=37.5$ Range, $181.7-162.5=19.2$

(B)

Height	Av. wt. for given ht.
185–189	82.5 } 71.9
180–184	71.3 }
175–179	66.4
170–174	62.8
165–169	59.2
160–164	57.9
155–159	54.2

Range, $185-160=25$ Range, $71.9-54.2=17.7$

Increase in average height . . . . $19.2 \div 6.55 = 2.93$
Corresponding increase in actual weight . . . . . . $37.5 \div 7.75 = 4.84$

$$\text{Ratio, } \frac{2.93}{4.84} = .60$$

Increase in average weight . . . . $17.7 \div 7.75 = 2.28$
Corresponding increase in height . . . . $25 \div 6.55 = 3.82$

$$\text{Ratio, } \frac{2.28}{3.82} = .60$$

Average height $=172.6$ cms. $\sigma_{\text{ht.}} = 6.55$ cms.
Average weight $=63.4$ kgs. $\sigma_{\text{wt.}} = 7.75$ kgs.

$$\text{Ratio, } \frac{\sigma_{\text{wt.}}}{\sigma_{\text{ht.}}} = \frac{7.75}{6.55} = 1.18$$

called a "scatter diagram," represents the paired heights and weights of 120 college men. The construction of such a scatter diagram is relatively simple. Along the left hand margin from bottom to top are laid off the steps of the height distribution; while along the top of the diagram from left to right are laid off the steps of the weight distribution. Each of the 120 men may now be located on the diagram with respect both to his height and his weight. Suppose, for example, that a man weighs 68 kgs. and is 176 cms. tall. His height locates him in the 3rd row from the top, and his weight in the 5th column from the left. Accordingly, this man belongs in the third "cell" of the 5th column and a tally is put in this cell. Note that in Diagram XVI there are 6 men and 6 tallies in this cell—that is, there are 6 men who weigh 65 to 69 kgs. and are 175 to 179 cms. tall. In the manner described every one of the 120 men has been located in some cell or square according to the two attributes, height and weight. Along the bottom of the diagram in the Fx row will be found the number of men who fall within each weight column (weight is the x-variable, page 60); while along the right hand margin in the Fy column are tabulated the number of men who fall within each height row (height is the Y-variable, page 60). Of course, both the Fy column and the Fx row total 120, the number of men in all. All of the frequencies in each cell may be totaled and written in numerical form as shown in the diagram. When only the total frequency in each cell is given, a scatter diagram becomes a correlation table (see Diagram XXI).

Several important facts may be gleaned from the scatter diagram as it stands. For example, we are able to classify all the men in a given weight-column with regard to height. In the 3rd column we find 28 men all of whom weigh 55 to 59 kgs. One of these 28 is 180 to 184 cms. tall; 4 are 175 to 179 cms. tall; 9 are 170 to 174 cms. tall; 7 are 165 to 169 cms. tall; and 7 are 160 to 164 cms. tall. In the same way we may classify all the men within any height-row accord-

ing to weight. In the row next to the bottom we find that of the 13 men who are 160 to 164 cms. tall, 1 weighs 45 to 49 kgs.; 2 weigh 50 to 54 kgs.; 7 weigh 55 to 59 kgs.; 1 weighs 60 to 64 kgs.; and 2 weigh 65 to 69 kgs. It is fairly clear, too, that the "drift" of paired heights and weights is from the upper right section of the diagram (the "high score" end) to the lower left hand section (the "low score" end). That is to say, even a superficial examination of the diagram indicates, in general, a fairly marked tendency for tall, medium, and short men to rank high, medium, and low, respectively, on the weight scale; and this observation holds, in spite of the scatter of heights or weights within any given "array" (an array is the distribution of cases within a given column or row). Without any further evidence, therefore, we should probably be willing to hazard the guess that the correlation between height and weight is positive and fairly high.

Suppose that we go a step further and calculate the average height of the men who weigh 45 to 49 kgs.—the men in column 1. The average height of these 3 men—using the guessed average method of Chapter I—is 162.5 cms., and this figure is entered at the bottom of the diagram. In the same way, we can find the *average height* of the men who fall in each of the succeeding weight-columns. These averages are tabulated under (*A*) and from the summary it is evident that for an actual weight increase of approximately 37.5 kgs.[1] (from 47.5 to 85) we have a corresponding increase in *average height* of 19.2 cms. (from 162.5 to 181.7). Thus it is clear that in our group of 120 college men, an increase of approximately 37.5 kgs. in *weight* is paralleled by increase of 19.2 cms. in *average height.*

Before going any further let us shift from height to weight, and applying the same method as above find the increase in *average weight* which corresponds to the *actual increase in height.* Taking the bottom row—the 3 men 155 to 159 cms. tall—we find that the average weight of this small group is

[1] The complete range is not taken into account because the data are scanty at the ends of the distribution.

54.2 kgs. The average weight of the 13 men who are 160 to 164 cms. tall is 57.9 kgs., and in like manner the average weight of each height-row may be found and entered in the "Average Weight" column. Summarizing the results for the group in (*B*) as we did in (*A*) above, we find that along with an increase in height of 25 cms. (160 to 185) there goes a corresponding increase in *average weight* of 17.7 kgs.[1] (71.9 to 54.2).

Now if the coefficient of correlation measures the mutual dependence or the degree of correspondence between two sets of scores or measures, we should expect the ratio $\frac{\text{increase in average height}}{\text{corresponding increase in weight}}$, e.g., $\frac{19.2}{37.5}$ to measure the correlation of height and weight, that is, to give us r. And likewise, and for the same reasons, we should expect the ratio $\frac{\text{increase in average weight}}{\text{corresponding increase in height}}$, e.g., $\frac{17.7}{25}$ also to measure the correlation. The two ratios work out, however, to be .51 and .71 respectively, which means evidently that neither is suitable as a measure of correlation, since the relation of height to weight should certainly be the same as the relation of weight to height in the same group.

The difficulty here—and while not an obvious one, it is easy to understand once it has been pointed out—is that we have failed to take account of the fact that the increases in height and weight, and naturally the ratios formed from them, depend for their numerical value upon the *units* which we have arbitrarily chosen for measuring height and weight. Thus while we have measured height in cms. and weight in kgs., it is clear that different units, say, of 1 mm. for height and 1 kg. for weight, or of 1 inch for height and 1 lb. for weight, would have given us very different ratios. In other words, the ratios which give the change in average height with corresponding change in weight, and the change in average weight with cor-

[1] The single F in the top row has been combined with the F of the row just below to prevent overweighting.

responding increase in height will vary according to the units in which height and weight are measured, and we have no way of telling which ratio (or what unit) is the right one. The best way out of this difficulty is to express the *changes* in height and weight in terms of the σ's of the *height* and *weight* distributions, respectively. It will make no difference then in what units our original measurements have been made, as changes in both height and weight will be recorded in terms of σ. The σ of the height distribution of our 120 men is 6.55 cms., and the σ of the weight distribution is 7.75 kgs. (see Diagram XVI). Accordingly, if we divide the increase in average height and the parallel increase in weight by 6.55 and 7.75 respectively, the ratio $\frac{\text{increase in average height}}{\text{corresponding increase in weight}}$ becomes $\frac{2.93}{4.84}$ or .605 (see Diagram XVI). And in like manner, if we divide the increase in average weight and the parallel increase in height by 7.75 and 6.55, respectively, the second ratio, $\frac{\text{increase in average weight}}{\text{corresponding increase in height}}$ becomes $\frac{2.28}{3.82}$ or .60. The two ratios are now equal, and either may be taken as representing the *coefficient of correlation*[1]—as giving the degree of association between height and weight in our group of 120 men.

This method of finding relationship is useful for demonstrating in a simple way what the ratio which we call the coefficient of correlation actually does. It is, however, neither a very practical nor precise method of finding a coefficient of correlation and is never used in actual practice. Its chief lack of precision lies in the fact that in estimating the range of scores or measures in either or both distributions (see footnote, page 155) we are often uncertain where to begin or end the series, due to the fact that the data are oftentimes scanty at the extremes of the distributions. As a matter of fact, the coefficient of correlation in the present problem was first found

[1] On a scale in which 1.00 denotes perfect relation.

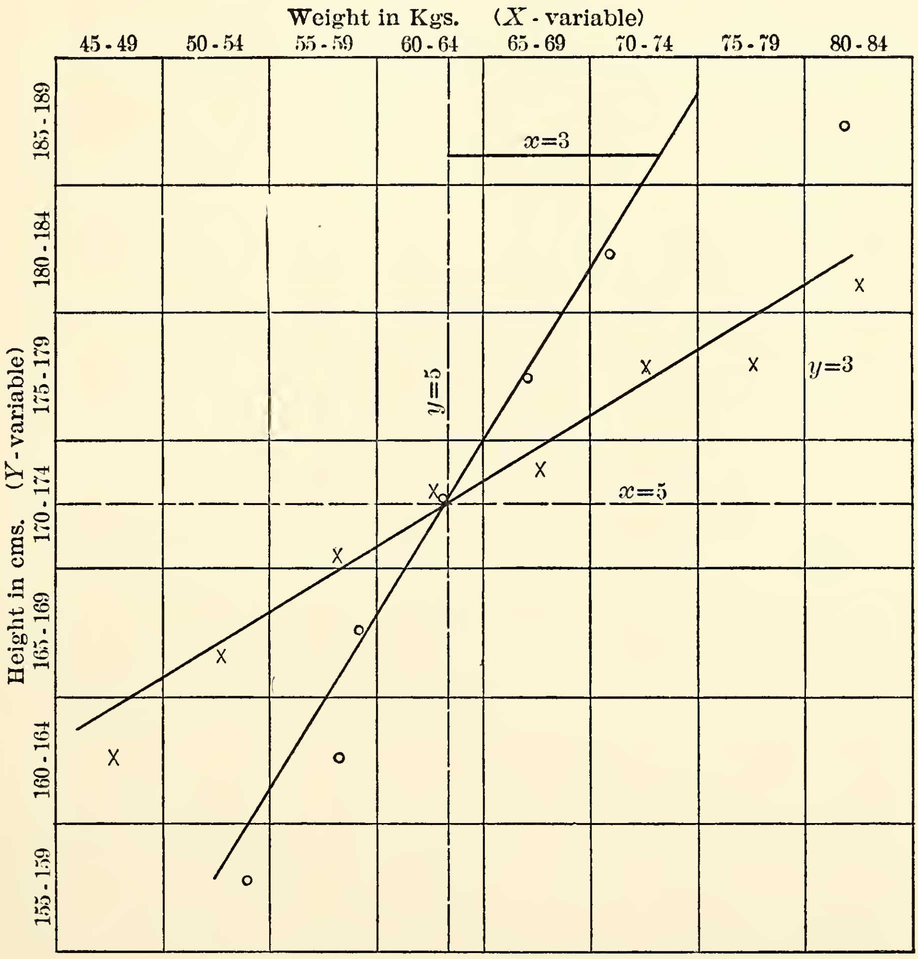
by the method given later on in Section III, and proper adjustment was then made in the ranges so as to give the correct r.

2. Graphical Representation of the Coefficient of Correlation

Not only can we represent the coefficient of correlation as a ratio, but we can also demonstrate *graphically* what a coefficient of correlation means. The correlation coefficient of .60 found in Diagram XVI between height and weight is shown graphically in Diagram XVII. In this diagram the distance taken to represent one unit (consider the step-interval as the unit) on the height scale and the distance taken to represent one unit on the weight scale have been selected with due regard for the difference in size of the two σ's in order that changes in height and weight may be comparable. This adjustment is a very simple one. We know from Diagram XVI that the $\sigma_{(\text{wt.})}$ which equals 7.75 kgs. is 1.18 times the $\sigma_{(\text{ht.})}$ which equals 6.55 cms. $\left(\text{since } \frac{7.75}{6.55} = 1.18\right)$. Hence it is only necessary that we take *each height-step* 1.18 times the length arbitrarily taken to represent *one weight-step*, in order that the X and Y distances may be comparable. (Since the weight distribution is laid off from left to right, and the height distribution from bottom to top, the first may be referred to as the X variable, and the second as the Y variable, see page 60.) To take a simpler case, if the σ for height were twice as large as the σ for weight, we should take each step on the height scale just $\frac{1}{2}$ each step on the weight scale.

When the diagram has been laid out in the manner described above represent by a cross the mean height of the men in each array—each weight column (these mean heights may be found from Diagram XVI). Next, draw a vertical line through the mean of the distribution of 120 weights, and a horizontal line through the mean of the distribution of 120 heights. [The average height of the 120 men is 172.6 cms., and their average weight is 63.4 kgs. (see Diagram XVI)]. With these two lines as coordinate axes, draw through their

intersection (the origin) a straight line which shall go through, or as close as possible to, each of the crosses which have been plotted. A rough—but fairly accurate—method of drawing



Average weight line drawn through 63.4 kgs.
" height " " " 172.6 cms.

DIAGRAM XVII

Coefficient of Correlation Shown Graphically

such a line is to stretch a black thread through the origin and shift it back and forth until it touches as many crosses as possible. The crosses at the extremes need not concern us very much, since they are located from only a few cases. This

sloping line, which may be called the line of "best fit," describes better than any other *straight* line the "run" of the crosses—the increase in average height which corresponds to the given increase in weight. Accordingly, to find the correlation simply find the ratio of the distance of *any point* on this sloping line from the horizontal or X-axis to the distance of the *same point* from the vertical or Y-axis. For example, if a convenient point P is taken with $x=5$ cms., its y distance (measured by mm. ruler) will be found to be approximately 3 cms., and the ratio $\frac{y}{x}$ is $\frac{3}{5}$ or .60. In like manner, the x and y coordinates of any other point on this sloping line will be found to give the ratio $\frac{y}{x}$ a value of .60.

Our sloping line pictures graphically the ratio $\frac{2.93}{4.84}$ —the correlation of .60—which we worked out in (1) above. This line, which will be known hereafter as the "regression line of height on weight," has important properties which will be considered later (page 173). Also in the following sections we shall give the equation of this line, which will enable us to draw it in on the diagram very much more accurately than can be done by the trial-and-error method described on page 159.

It is a comparatively easy though not a necessary task to verify the correlation coefficient of .60 found from the regression line of height on weight by drawing in the second "regression line," that of weight on height. This can be done by designating the means of the different height-rows by circles in exactly the same manner in which we marked the means of the weight-columns by crosses. (The means of the rows may be obtained from Diagram XVI.) The mean of the lowest row is 54.2, of next above 57.9, etc. When all of the circles have been correctly placed, we draw a straight line which shall go through—or as close as possible to—each circle, just as we did with the crosses above. Now if a point P' is taken on this second line with a $y=5$ cms., its x distance will be found to be

approximately 3 cms., and the ratio $\frac{x}{y}$ is .60. This relation holds for any point on the line. Both regression lines, therefore, give us the same measure of the correlation between height and weight.

Diagram XVII is still further useful in showing just what a correlation of 1.00, 0, or −1.00 is graphically. Suppose (1) that the two regression lines in the figure move together until they coincide in such a way as to make an angle of 45 degrees with the horizontal or X-axis. The x value of any point on this "compound" line will always equal its y value—hence the ratios $\frac{y}{x}$ and $\frac{x}{y}$ are always equal to each other [1] and r equals 1.00 (see Diagram XVIII). Accordingly, in *perfect positive* correlation, all the crosses and all the circles in a correlation diagram fall along a single straight line which runs from the upper right hand section of the diagram (the 1st quadrant) to the lower left hand section (the 3rd quadrant). The tallest man is the heaviest, the next tallest, the next heaviest, and throughout the entire 120 the correspondence of height and weight is always 1 to 1.

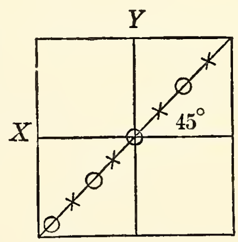


DIAGRAM XVIII

Now suppose (2) that the first regression line, the line through the means of the height arrays in the columns—through the crosses—moves around until it coincides with the X-axis, the line through the average of *all the heights* in the table. And suppose again that the second regression line, the line through the means of the weight arrays in the rows—through the circles—moves around until it coincides with the Y-axis, the line through the average of *all the weights* in the table. The ratios $\frac{y}{x}$ and $\frac{x}{y}$ are now both equal to 0 (since in the first case x, and in the second case y, equals 0) and r, the

[1] This is true also because the compound regression line becomes the diagonal of a square. Again, the tangent of an angle of $45° = 1.00$.

coefficient of correlation, equals 0. The conclusion that $r=0$ might also be drawn from the fact that under the conditions described the *average height* is the same for the whole range of weights and the *average weight* the same for the whole range of heights. Hence, a man of average height is equally liable to be heavy, medium, or light, and a man of average weight equally liable to be tall, medium, or short. (Compare with the case in which the average tapping rate was the same for very high, high, and medium high Alpha scores, page 150.) A picture of zero correlation is shown in Diagram XIX.

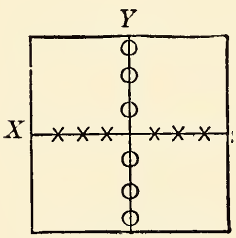


DIAGRAM XIX

Lastly, suppose (3) that the two regression lines swing around until they run from the upper left hand section (the 2nd quadrant) to the lower right hand section (the fourth quadrant). Now if the two lines again coincide so as to make an angle of 45 degrees with the X-axis—as described in (1)—the x of any point on this compound line will always equal the y of the same point, and the ratios $\frac{y}{x}$ and $\frac{x}{y}$ will again always equal 1.00. A glance at the figure will show, however, that either the x or the y of these ratios must always be negative, and for this reason the ratios will always be negative. The coefficient of correlation, therefore, equals -1.00, and the relation is perfect but inverse. In *perfect negative* correlation, it is clear then that all of the crosses and all of the circles fall along a single straight line which runs from the upper left to the lower right hand corner of the diagram. The tallest man in the group is the lightest, the next tallest the next lightest, and as height decreases weight increases progressively. (Diagram XX.)

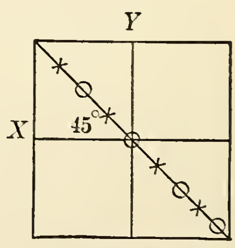


DIAGRAM XX

The regression lines coincide only when the correlation is perfect—positive or negative. For degrees of correlation

between these limits, the two regression lines are separate, and take intermediate positions as shown in Diagram XVII for an $r = .60$.

III. The Calculation of the Coefficient of Correlation by the Product-Moment Method

1. The Product-Moment Formula When Deviations Are Taken from the Guessed Averages of the Two Distributions

With the meaning of a coefficient of correlation firmly in mind as a result of the discussion of the last section, we are now ready to consider the calculation of r by the product-moment method.[1] Diagram XXI will serve as an illustration of the computations involved. This correlation table gives the paired heights and weights of 120 college men and is derived from the scatter diagram for the same data shown in Diagram XVI. The complete process of calculating r is outlined in the following steps. (Diagram XXI should be constantly referred to in the discussion that follows.)

Step I

Construct a scatter diagram and from it a correlation table as described on page 154.

Step II

Guess an average for the height distribution (given in the F_y column), and draw double lines to mark off the *row* which contains the $GA_{(\text{ht.})}$, as shown in Diagram XXI. Note that the average for the height distribution has been guessed at 172.5 (midpoint of interval 170–174) and that D_y's have been taken from this point. Now fill in the FD_y and the FD_y^2 columns. From the first column the correction C_y (c_y in units of step) is obtained; and this correction together with the sum of the FD_y^2 column will give the σ of the height distribution, σ_y. The value of σ_y is 6.55 cms. (1.31×5)—see calculations in the Diagram.

[1] The r found by this method is often called the "Pearson r" after Prof. Karl Pearson, who devised the product-moment formula, following Bravais's earlier work.

DIAGRAM XXI

Calculation of the Product-Moment Coefficient of Correlation between the Heights and Weights of 120 College Men

Weight in kgs. (X variable)

Height in cms. (Y-variable)	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	Fy	D_y	FD_y	FD_y^2	$\Sigma x'y'$ +	$\Sigma x'y'$ −
185-189								(12) 1 12	1	3	3	9	12	
180-184			(-2) 1 -2	3 0	(2) 3 6	(4) 4 16	(6) 2 12	(8) 3 24	16	2	32	64	58	2
175-179			(-1) 4 -4	11 0	(1) 6 6	(2) 3 6	(3) 2 6	(4) 2 8	28	1	28 (63)	28	26	4
170-174		2 0	9 0	11 0	8 0	2 0	1 0		33	0				
165-169	(3) 1 3	(2) 5 10	(1) 7 7	10 0	(-1) 3 -3				26	−1	−26	26	20	3
160-164	(6) 1 6	(4) 2 8	(2) 7 14	1 0	(-2) 2 -4				13	−2	−26	52	28	4
155-159	(9) 1 9	(6) 1 6		1 0					3	−3	− 9 (−61)	27	15	
Fx	3	10	28	37	22	9	5	6	120		2	206	159	−13
													(146)	
Dx	−3	−2	−1	0	1	2	3	4						
FDx	−9	−20	−28 (−57)		22	18	15	24 (79) = 22						
FD_x^2	27	40	28		22	36	45	96 = 294						

$$c_y = \frac{2}{120} = .017 \qquad c_x = \frac{22}{120} = .183$$

$$c^2_y = .0003 \qquad c^2_x = .0334$$

$$C_y = .085 \qquad C_x = .915$$

$$\sigma_y = \sqrt{\frac{206}{120} - .0003} \times 5 \qquad \sigma_x = \sqrt{\frac{294}{120} - .0334} \times 5$$

$$\sigma_y = 1.31 \times 5 \qquad \sigma_x = 1.55 \times 5$$

$$\sigma_y = 6.55 \qquad \sigma_x = 7.75$$

Calculation of r:

$$r = \frac{\frac{146}{120} - .017 \times .183}{1.31 \times 1.55}$$

$$r = .60$$

$$PE_r = \frac{.6745[1-(.60)^2]}{\sqrt{120}}$$

$PE_r = .04$ (Table XVIII)

Now guess an average for the weight distribution (given in the F_x row) and draw double lines to designate the column which contains the $GA_{(\text{wt.})}$. The average of the weight distribution has been guessed at 62.5 (midpoint of interval 60–64) and D_x's have been taken from this point. Fill in the FD_x and FD_x^2 rows. From these rows the correction C_x

(c_x units of step) and the σ of the weight distribution σ_x, may be obtained. The value of σ_x is 7.75 kgs. (1.55×5)—see calculations on the Diagram.

Step III

The calculations in Step II simply repeat the familiar process of finding a σ by the Guessed Average Method. (Chapter I, page 35.) Our first *new* task is to fill in the $\Sigma x'y'$ column. The entries in this column may be either $+$ or $-$, and hence two columns are provided under $\Sigma x'y'$, one for plus and one for minus entries.

The procedure for determining the entries in the $\Sigma x'y'$ column may be illustrated by taking the single entry in the only occupied cell in the topmost row. The deviation of this cell from the *GA* of the weight distribution, that is, its D_x, is 4 steps, and its deviation from the *GA* of the height distribution, its D_y, is 3 steps. Hence, the product of the deviations of this *cell*—its "product-moment"—from the two guessed averages is 4×3 or 12, and a small figure 12 is placed in the upper right hand corner of the cell.[1] Moreover, since the "product-moment" of the 1 frequency in this cell is $1(4 \times 3)$ or 12 also, a figure 12 is placed in the *lower left* hand corner of the cell to denote the product of the deviations (or the *product-deviation*) of this *single frequency* from the two *GA*'s. There are no other frequencies in the cells of this row, and 12 is placed at once in the $\Sigma x'y'$ column[2] under the $+$ sign.

Now let us consider the next row from the top, taking the cells in order from right to left. The cell below the one whose product-deviation we have just found, also deviates 4 steps from the *GA* of the weight distribution (its $D_x = 4$) but its deviation from the *GA* of the height distribution is only 2 steps

[1] We may take the coordinates of this cell to be $x = 4$, and $y = 3$. The first is obtained by counting over 4 steps from the vertical column containing the *GA* for weight, and the second by counting up 3 steps from the horizontal row containing the *GA* for height. In each case the unit of measurement is the step-interval.

[2] The prime (′) of x and y deviations is to indicate that all deviations are taken from the two *GA*'s.

(its $D_y=2$). Hence the product-deviation of this cell is 4×2 or 8 [note the small (8) in the upper right hand corner of the cell], and since there are 3 frequencies in the cell, each with a product-deviation of 8, the final entry in the lower left hand corner of this cell is $3(4\times2)$ or 24. In like manner, the product-deviation of the 2nd cell in the row is 6,—its $D_x=3$, and its $D_y=2$,—and since there are 2 frequencies in the cell, the final entry is $2(3\times2)$ or 12. Each of the 4 frequencies in the third cell has a product-deviation of 4 (the D_x of the cell is 2, and the D_y is 2 also) and the final cell entry is $4(2\times2)$ or 16. In the 4th cell each of the 3 frequencies has a D_x of 1 and a D_y of 2, and the product deviation is $3(1\times2)$ or 6. The entry of the 5th cell, the cell in the $GA_{(\text{wt.})}$ column, is 0, since $D_x=0$, and of course $3(2\times0)=0$. Notice particularly the entry in the last cell of this row, viz., -2. This negative entry results from the fact that the deviation of this cell from the $GA_{(\text{wt.})}$, its D_x, is -1, and its D_y is 2; the product-deviation of its single frequency, therefore, is $1(-1\times2)$ or -2. Now total separately the plus and minus $x'y'$'s in this row. The results, 58 and -2, are entered separately in the $\Sigma x'y'$ column under the appropriate signs.

The final entries of the cells in the other rows in the table and the sums of the product-deviations of each row are obtained in the manner described above. It must be borne in mind in calculating $x'y'$'s that the product-deviations of *all* frequencies in the *first* and *third* quadrants are positive, while the product-deviations of *all* the frequencies in the *second* and *fourth* quadrants are negative (see page 162). Also remember that all frequencies in either the column containing the $GA_{(\text{wt.})}$ or in the row containing the $GA_{(\text{ht.})}$ have 0 product-deviations, since in one case the D_x, and in the other the D_y, equals 0.

All frequencies in any given row have the same D_y, and for this reason the arithmetic of calculation may be considerably reduced if each frequency in the row is first multiplied by its D_x, and the sum of these deviations multiplied once for all by the common D_y. To illustrate, for the 2nd row from the

bottom—taking the cells from right to left—when we multiply the frequency of each cell by its D_x, the result is $(2\times1)+(1\times0)+(7\times-1)+(2\times-2)+(1\times-3)$ or -12. Now multiplying this partial "deviation-sum" by the D_y of the *whole row*, i.e., by -2, we get 24 at the final $\Sigma x'y'$ entry for the row. This result checks the 28 and -4 entered separately in the $\Sigma x'y'$ column. This shorter method is useful in getting the total $\Sigma x'y'$ entry of a given row quickly. It is less easy to check for errors, however, than the method of getting the entry for each cell separately, illustrated on page 166.[1]

Step IV

When the sum of the product-deviations of each row have been entered in the $\Sigma x'y'$ column, the algebraic sum of the $\Sigma x'y'$ column may be obtained (e.g., $159-13=146$). The coefficient of correlation is then found by the formula:

$$r=\frac{\dfrac{\Sigma x'y'}{N}-c_xc_y}{\sigma_x\sigma_y}, \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (23)$$

Substituting for $\dfrac{\Sigma x'y'}{N}$, $\dfrac{146}{120}$; for c_x, .183; for c_y, .017; and for σ_x and σ_y, 1.55 and 1.31, respectively, (see Diagram XXI for figures) r is found to equal .60.

Notice that the terms c_x, c_y, σ_x and σ_y are all left in units of step-interval when substituted in formula (23). This is done simply because all product-deviations ($x'y'$'s) are in step-units and hence it is very much easier to keep *all* the other terms in the formula, and in consequence both numerator and denominator, in step-units. By this procedure the value of the

[1] Printed charts for facilitating the calculation of coefficients of correlation by the product-moment method are now available. Examples are the Ruch-Stoddard Correlation Charts, University Bookstore, Iowa City, Iowa, and Thurstone Correlation Data Sheet, C. H. Stoelting & Co., Chicago. The first of these gives the product-deviation of each cell printed on the chart. Otis has also devised a correlation chart based on the product-moment method which does away with the necessity of finding the $x'y'$'s. This chart is published with directions for its use by the World Book Co., Yonkers, N. Y.

fraction—the coefficient of correlation—is not changed and the arithmetic is considerably reduced.

2. The Product-Moment Formula When Deviations Are Taken from the Actual Averages of the Two Distributions

Since formula (23) assumes that all x and y deviations have been taken from the two guessed averages, for this reason it is necessary to correct $\frac{\Sigma x'y'}{N}$ by the amount of the two corrections, c_x and c_y. If deviations are taken from the actual averages of the two distributions instead of from the GA's, no correction is needed, as both c_x and c_y then equal 0. Thus when deviations are taken from the two averages, formula (23) becomes

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y}, \quad \ldots \ldots \ldots \quad (24)$$

and this is the form in which the product-moment formula is usually written. The formula may be put in still another form. If we write $\sqrt{\frac{\Sigma x^2}{N}}$ for σ_x and $\sqrt{\frac{\Sigma y^2}{N}}$ for σ_y, the formula then becomes (the Ns cancel)

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2} \cdot \sqrt{\Sigma y^2}}, \quad \ldots \ldots \quad (25)$$

in which the x and y deviations are from the averages as in (24) and $\sqrt{\Sigma x^2}$ and $\sqrt{\Sigma y^2}$ are the sums of the squared deviations from the two averages.

Formula (23) should always be used when there are more than, say, 30 or 40 cases. Formula (25) may be used, to advantage, however, with short series when the purpose of the experimenter is to find whether there is any relation present rather than to discover the degree of relation very accurately. No correlation table is required with formula (25). An illustration of the use of this formula is given in Table XVII, in which the problem is to find the correlation between the scores

TABLE XVII

To Illustrate the Calculation of r when Deviations are Taken from the Averages of the Distributions

Individual	Score in Test 1(X)	Score in Test 2(Y)	x	y	x^2	y^2	xy
A	50	22	−12	−8.4	144	70.56	100.8
B	53	25	− 9	−5.4	81	29.16	48.6
C	56	34	− 6	3.6	36	12.96	−21.6
D	58	28	− 4	−2.4	16	5.76	9.6
E	60	26	− 2	−4.4	4	19.36	8.8
F	61	30	− 1	− .4	1	.16	.4
G	61	32	− 1	1.6	1	2.56	− 1.6
H	64	30	2	− .4	4	.16	− .8
I	67	28	5	−2.4	25	5.76	−12.0
J	70	34	8	3.6	64	12.96	28.8
K	71	36	9	5.6	81	31.36	50.4
L	73	40	11	9.6	121	92.16	105.6
Average	62	30.4			578	282.92	317.0

Average (Test 1) = 62.0

Average (Test 2) = 30.4

$$r = \frac{317}{\sqrt{578} \cdot \sqrt{282.92}} = .78$$

$$PE_y = \frac{.6745(1-(.78)^2)}{\sqrt{12}} = .08$$

made on two tests of association by 12 adults. The steps in finding r may be outlined as follows:

Step I

Find the average of Test 1 and the average of Test 2. In the table the first average is 62.0, and the second, 30.4.

Step II

Find the deviations of each score in Test 1 from its average, 62, and enter in column x. (The deviations from the average of the first test may be called x-deviations, those from the average of the second test, y-deviations.) Find the deviation of each score in Test 2 from its average, 30.4, and enter in column y.

Step III

Square all x-deviations, and all y-deviations, and enter these squares in columns x^2 and y^2, respectively.

Step IV

Multiply the corresponding x and y deviations and enter these products in the xy column.

Step V

Substitute for Σxy (317), for Σx^2 (578), for Σy^2 (282.92) in formula (25) as shown in Table XVII, and solve for r.

IV. The Probable Error of a Coefficient of Correlation

The PE of an r may be found from the formula,

$$PE_r = \frac{.6745 \times (1 - r^2)}{\sqrt{N}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

If we substitute in formula (26) the $r = .60$ and the $N = 120$ of the height-weight problem (see Diagram XXI), PE_r will equal .04.[1] This means that the chances are even that the "true" r falls within the limits $.60 \pm .04$, or between .56 and .64; and that the chances are 9930 in 10,000 (Table XI) that the true r falls within the limits $.60 \pm 4 \times .04$, or between .44 and .76. By the true r is meant (see page 118) that r which we should expect to get between height and weight in the population from which our group of 120 is, presumably, a random sampling.

To be reasonably sure that there is some correlation present an obtained r should be at least 4 times its PE. For example, given the situation in which r is exactly 4 times its PE, in which, say, $r = .16$ and $PE_r = .04$, we can only be *sure* that the true r falls within the limits $.16 \pm 4 \times .04$, or between 0 and .32. It is customary, therefore, not to consider an r as reliable—as indicative of a correlation at least better than 0—unless it is at least 4 times its PE. To be certain of a low degree of correlation an r should be 5 or 6 times its PE.

We found in Chapter III that the reliability of the difference between two averages or two medians can be calculated by

[1] If we know r and N, the PE_r may be read directly or by interpolation from Table XVIII.

means of the formulas for $\sigma_{(diff.)}$ and $PE_{(diff.)}$ (see page 128). In the same way, the reliability of the difference between two obtained r's can be found from the size of the PE of their difference.

TABLE XVIII

Probable Errors of the Coefficient or Correlation for Various Numbers of Measures (N) and for Various Values of r

Number of Measures	0.0	0.1	0.2	0.3	0.4	0.5	0.6
	Correlation Coefficient r						
20	1508	1493	1448	1373	1267	1131	0965
30	1231	1219	1182	1121	1035	0924	0788
40	1067	1056	1024	0971	0896	0800	0683
50	0954	0944	0915	0868	0801	0715	0610
70	0806	0798	0774	0734	0677	0605	0516
100	0674	0668	0648	0614	0567	0506	0432
150	0551	0546	0529	0501	0463	0413	0352
200	0477	0472	0458	0434	0401	0358	0305
250	0426	0421	0409	0387	0358	0319	0272
300	0389	0386	0374	0354	0327	0292	0249
400	0337	0334	0324	0307	0283	0253	0216
500	0302	0299	0290	0274	0253	0226	0193
1000	0213	0211	0205	0194	0179	0160	0137
Number of Measures	0.65	0.7	0.75	0.8	0.85	0.9	0.95
20	0871	0769	0660	0543	0419	0287	0147
30	0711	0628	0539	0444	0342	0234	0120
40	0616	0544	0467	0384	0296	0203	0104
50	0551	0486	0417	0343	0265	0181	0093
70	0466	0411	0353	0290	0224	0153	0079
100	0391	0345	0294	0242	0187	0128	0066
150	0318	0281	0241	0198	0153	0105	0054
200	0275	0243	0209	0172	0133	0091	0047
250	0246	0218	0187	0154	0118	0081	0042
300	0225	0199	0170	0140	0108	0074	0038
400	0195	0172	0148	0122	0094	0064	0033
500	0174	0154	0132	0109	0084	0057	0029
1000	0123	0109	0093	0077	0059	0041	0021

The formula for $PE_{(diff.)}$ between two r's is

$$PE_{(diff.\ r_1 - r_2)} = \sqrt{PE^2_{r_1} + PE^2_{r_2}}, \quad . \quad . \quad . \quad . \quad (27)$$

in which PE_{r_1} and PE_{r_2} are the PE's of the two r's to be compared, and must first be obtained from formula (26).

The value of formula (27) may be illustrated by the following problem. Suppose that in a group of 100 eight year old boys the

r between IQ and the A-cancellation test is .20 with a PE of .065; and that in a group of 110 eight year old girls the r between the same two tests is .25 with a PE of .06. The correlation is .05 higher for girls than for boys. Is this difference sufficiently large to indicate that the true correlation between IQ and the A-test is higher for 8 year old girls than for 8 year old boys? To answer this question, we must determine the PE of the difference between the two r's. From formula (27), $PE_{(\text{diff.}\,r_1-r_2)} = \sqrt{(.065)^2+(.06)^2} = .09$, and comparing the obtained difference of .05 with the $PE_{(\text{diff.})}$, we find that $\frac{D}{PE_{(\text{diff.})}} = .556$. This means (see Table XV) that there are only 64 chances in 100 of a real difference, a difference greater than 0, between the true correlations of IQ and the A-test for 8 year old boys and girls. The difference of .05 is, therefore, quite unreliable. To be completely reliable the obtained difference should be at least $4 \times .09$ or .36. (A difference is considered reliable when $\frac{D}{PE_{(\text{diff.})}}$ is 4 or more, see page 133.) In the present case the obtained difference is only about 14 per cent of what it should be in order to guarantee a true difference between the r's of the boys and girls.

The formulas for PE_r and $PE_{(\text{diff.}\,r_1-r_2)}$ are subject to the same restrictions and must be interpreted with the same caution as the other standard and probable error formulas (see Chapter III, page 145). In order to be of any real value as measures of reliability, PE_r and $PE_{(\text{diff.})}$ should be calculated for r's obtained from random and reasonably large samples. PE's found for r's obtained from small and obviously selected groups may give an entirely false picture of the observed coefficient's reliability—especially when the coefficient is large. An r of .90 found from 20 cases, for instance, is unreliable despite the fact that $PE_r = .03$ (see Table XVIII). Another sample of 20 drawn from the same population might give an r one half as large.

V. The Regression Equations

1. The Regression Equations in Deviation Form

We have already discovered (Diagram XVII) that there are two regression lines in a correlation table, and that the first "best fits" the means of the successive columns (the average heights, represented by crosses) while the second "best fits" the means of the rows (the average weights, represented by circles). These lines of "best fit" were seen to be of value in showing graphically the change in average height accompanying a given change in weight, and the change in average weight accompanying a given change in height. Moreover, we found that either line will measure the correlation directly when the x and y steps in the diagram have been laid out with due allowance for the difference in size of the σ's of the X and Y distributions.

This last use of the regression line is of little practical value, however. It is very much easier to draw up a correlation table without bothering about the difference in the two σ's, and find r by the product-moment formula as shown in Diagram XXI, than to try and estimate r from the regression lines. In fact, the real value of the regression lines is not to give r, but to enable us to "predict" an individual's "most probable" standing in a test or series of measures, given his standing in another test or series of measures.

We may describe briefly how this is done. Suppose that we wish to estimate a man's height from our correlation table, knowing his weight to be 68 kgs. Now the best possible "guess" that we can make of this man's height is to give the average height of all men who fall in the 65–69 weight interval. From Diagram XVI the "mean weight" of the 25 men in this column is found to be 173.6 cms., and hence 173.6 cms. is the most likely height of a man who weighs 68 kgs. In like manner, the most probable height of a man who weighs 72 kgs. is 178.6 cms.—the mean height of the 9 men who fall in the weight column 70–74 kgs. In general, then, the most probable height

of any man is the *mean* of the heights of all the men in the group who weigh the same (approximately) as he—who fall in the same weight column.[1] The line which best fits the mean heights of the successive weight-columns is the line which gives the change in average height with the change in weight (the line through the crosses in Diagram XVII). Given a man's weight, therefore, we can best "predict" his height from the regression line of height on weight; and by analogy, given a man's height, we can best predict his weight from the regression line of weight on height (the line through the circles in Diagram XVII).

If we had the equations of the two regression lines, it would seem obvious that estimates could be made from these much more efficiently and quickly than from the plotted regression lines. For then knowing a man's standing in the X-variable (his weight) we should be able on substituting in the equation connecting X and Y to find directly his most probable standing in the Y-variable (height). The equations of the two regression lines have been deduced by Prof. Karl Pearson, who took as his criterion the idea of the "best fitting" line. Pearson's method, briefly, was to find the equation of that line from which the sum of the squares of the deviations of the means in the different arrays (the rows or the columns) is the least possible.[2] There are, of course, two such lines. The one "best fits" the means of the rows, the other "best fits" the means of the columns.

The equation of the line drawn through the means of the columns (the crosses in Diagram XVII) is written in its simplest form[3] as

$$y = r \cdot \frac{\sigma_y}{\sigma_x} \cdot x. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (28)$$

[1] There is a certain error of estimate made in taking a man's most probable height as being the average of his weight-group. The method of finding the size of this error will be considered later on page 183.

[2] For a mathematical treatment of the application of the Method of Least Squares to the problem of deducing the regression equations, see Jones, *A First Course in Statistics*, 1921, pp. 106ff and 271.

[3] A brief review of the equation of a straight line and of the method of plot-

The expression $r \cdot \frac{\sigma_y}{\sigma_x}$ is called the *regression coefficient* and is often replaced in the equation by the expression b_{yx} or b_{12}, so that (28) is sometimes written $y = b_{yx} \cdot x$ and $y = b_{12} \cdot x$.

If we substitute the values of r, σ_y, and σ_x,—obtained from Diagram XXI—in formula (28) we have

$$y = .60 \times \frac{6.55}{7.75} \cdot x \text{ or } y = .51x,$$

as the equation which measures the regression of height on

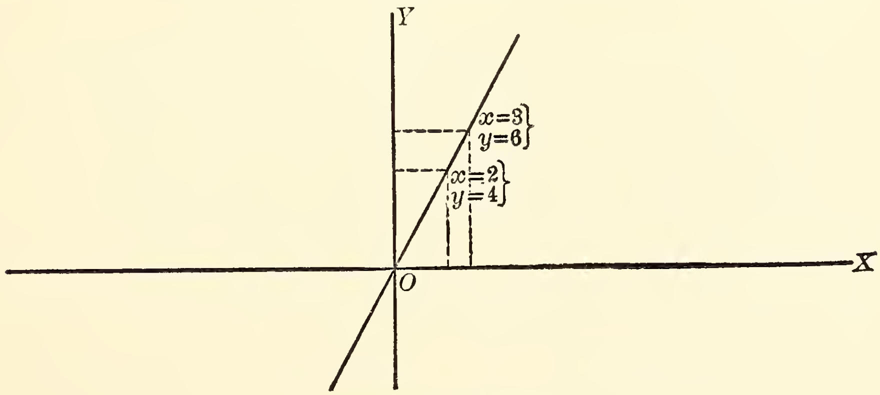


DIAGRAM XXII

ting a simple linear equation is given in order to simplify the discussion of the regression equations.

Let X and Y be coordinate axes, or axes of reference. Now suppose that we are given the equation $y = 2x$ and are required to represent the relation between x and y graphically. To do this we substitute values for x in the equation and compute the corresponding values of y. When $x = 2$, for example, $y = 2 \times 2$ or 4; when $x = 3$, $y = 2 \times 3$ or 6. In like manner, given any x value, we can compute the y which will " satisfy " the equation, that is, make the left side equal to the right. Now if the series of points determined from the pairs of x and y values as given by the equation are plotted with respect to the X and Y axes (see Diagram XXII) they will be found to fall along a straight line, and this straight line will picture the relation of x and y, $y = 2x$. This line will pass through the origin, since when $x = 0$, y also equals 0. The equation $y = 2x$ represents, then, a straight line which passes through the origin and the relation of its points is such that $\frac{y}{x}$ (called the slope of the line) always equals 2.

The general equation of any straight line which passes through the origin may be written $y = mx$, where m is the slope of the line. If we replace the m of the general formula by the expression $r \cdot \frac{\sigma_y}{\sigma_x}$ we see at once that the regression equation in *deviation form* is simply the equation of a straight line which goes through the origin.

weight. This equation represents a straight line through the origin, and hence it is a simple matter to plot it, as shown in Diagram XXIII. First, however, we must draw a vertical line through the point 63.4 kgs., the mean of all the weights (the X's) in the table, and a horizontal line through 172.6 cms., the mean of all the heights (the Y's) in the table. These two lines are the coordinate axes. Now since our plotted line must go through the origin [see note(3), page 175], only one other point is needed to determine it. If $x=2$ (any value will do just as well), y becomes $.51\times 2$ or 1.02. To plot this point, measure out 2 units from the origin along the horizontal axis and go up 1.02 units from the same line. This will locate the point, $x=2$, $y=1.02$. (Any convenient scale may be used for measuring off x and y distances—a mm. rule is useful.)

The line drawn through the point just located and the origin (0, 0) is the regression line of height on weight. From the equation, it is clear that a point on this line with an x-value of 1.00 has a corresponding y-value of .51 (substitute $x=1$ in the equation and $y=.51$). This means that a deviation of 1 unit from the mean of the X's (from the vertical line drawn through the mean weight of the group) is accompanied by just .51 time as much deviation from the mean of the Y's (from the horizontal line drawn through the mean height of the group) (see Diagram XXIII). Put concretely, a man who stands 1 kg. *above* the average weight of the group is most probably .51 cm. *above* the mean height of the group also—if his weight is 64.4 kgs. (63.4+1.00) his height is probably 173.11 cms. (172.6+.51). To take another example, the man who weighs 60 kgs.—stands 3.4 kgs. *below* the mean weight—is most probably 170.87 cms. tall—stands 1.73 cms. *below* the mean height. In this example, we substitute $x=-3.4$ in the equation, and $y=-1.73$. In general then we know from the regression equation that the most probable deviation of any individual in our group[1] from the mean

[1] Or in the population from which our group of 120 is drawn, provided the group is a random sample.

DIAGRAM XXIII

Illustrating Position of the Regression Lines, and Calculation of the Regression Equations

(Calculation of r repeated from Diagram XXI)

Weight in kgs. (X-variable)

Height in cms. (Y-variable)	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	F_y	D_y	FD_y	FD_y^2	$\Sigma x'y'$ +	$\Sigma x'y'$ −
185-189								1 (12) 12	1	3	3	9	12	
180-184			1 (−2) −2	3 (0)	3 (2) 6	4 (4) 16	2 (6) 12	3 (8) 24	16	2	32	64	58	2
175-179			4 (−1) −4	11 (0)	6 (1) 6	3 (2) 6	2 (3) 6	2 (4) 8	28	1	28 (63)	28	26	4
170-174		2 (0)	9 (0)	11 (0)	8 (0)	2 (0)	1 (0)		33	0				
165-169	1 (3) 3	5 (2) 10	7 (1) 7	10 (0)	3 (−1) −3				26	−1	−26	26	20	3
160-164	1 (6) 6	2 (4) 8	7 (2) 14	1 (0)	2 (−2) −4				13	−2	−26	52	28	4
155-159	1 (9) 9	1 (6) 6		1 (0)					3	−3	− 9 (−61)	27	15	
F_x	3	10	28	37	22	9	5	6	120		2	206	159 (146)	−13
D_x	−3	− 2	− 1	0	1	2	3	4						
FD_x	−9	−20	− 28 (−57)		22	18	15	24 (79) = 22						
FD_x^2	27	40	28		22	36	45	96 = 294						

$$cy=\frac{2}{120}=.017 \qquad cx=\frac{22}{120}=.183$$

$$c^2y=.0003 \qquad c^2x=.0334$$

$$Cy=.085 \qquad Cx=.915$$

$$GA(Y)=172.5 \qquad GA(X)=62.5$$

$$\text{Aver.}(Y)=172.6 \qquad \text{Aver.}(X)=63.4$$

$$\sigma_y=\sqrt{\frac{206}{120}-.0003}\times5=6.55 \qquad \sigma_x=\sqrt{\frac{294}{120}-.0334}\times5=7.75$$

Calculation of r:

$$r=\frac{\frac{146}{120}-.017\times.183}{1.31\times1.55}=.60$$

$$PE_r=.04$$

Calculation of Regression Equations:

I. Deviation Form:

(1) $y=.60\times\frac{6.55}{7.75}x=.51x$

(2) $x=.60\times\frac{7.75}{6.55}y=.71y$

II. Score Form:

(1) $Y-172.6=.51(X-63.4)$
$Y=.51X+140.3$

(2) $X-63.4=.71(Y-172.6)$
$X=.71Y-59.1$

Calculation of Standard Errors of Estimate:

$\sigma_{(\text{est. }Y)}=6.55\times.8=5.2$ cms.
$\sigma_{(\text{est. }X)}=7.75\times.8=6.20$ kgs.

height is just .51 as great as his deviation from the mean height. Hence, given a man's *deviation from the mean weight*, we are able to predict his *most probable deviation from the mean height* of the group.

The regression equation, $y = r \cdot \frac{\sigma_y}{\sigma_x} \cdot x$, is known as the regression equation of Y on X in *Deviation Form.* Stated generally, this equation measures the *most probable deviation* of any Y measure from the mean Y corresponding to a *known deviation* in the X measure from the mean X.

The equation of the second regression line drawn through the means of the rows (the circles of Diagram XVII) is written

$$x = r \cdot \frac{\sigma_x}{\sigma_y} \cdot y. \quad \ldots\ldots\ldots \quad (29)$$

This equation measures the regression of X on Y and in the present problem, of weight on height. The regression coefficient $r \cdot \frac{\sigma_x}{\sigma_y}$ is sometimes replaced by the expression b_{xy} or b_{21}, so that (29) is often written $x = b_{xy} \cdot y$ or $x = b_{21} \cdot y$.

If we substitute in (29) the values of r, σ_x, and σ_y found from Diagram XXI, we have

$$x = .60 \times \frac{7.75}{6.55} \cdot y \text{ or } x = .71y,$$

as the equation which measures the regression of weight on height. This equation, like the other, represents a straight line through the origin; and consequently, one point on the line together with the origin (0, 0) are sufficient to plot the line. Put $y = 1$ in the equation, and x will equal .71. Now plot the point $x = .71$, $y = 1.00$ on the diagram, and draw the regression line through this point and the origin (see Diagram XXIII).

It is evident from the second regression equation that a deviation of 1 cm. from the mean of all the heights (Y's) is most probably accompanied by a deviation of .71 kg. from the

mean of all the weights (X's); or put in a different way, the most probable deviation of any man from the mean weight is just .71 as great as his deviation from the mean height. A man 180 cms. tall, for example (7.4 cms. above the mean height), most probably weighs 68.65 kgs.—is 5.25 kgs. above the mean weight). (To get this result substitute 7.4 for y in the equation, and solve for x.)

The equation $x=r\cdot\frac{\sigma_x}{\sigma_y}\cdot y$ is known as the regression equation of X on Y in *deviation form*. To summarize briefly it measures the *probable deviation* of an X-measure from the average X, corresponding to a *known deviation* in the Y-measure from the average Y.

Although there are two regression equations, both of which involve x and y, the student must bear in mind the important fact that the two equations cannot be used interchangeably and that neither can be used to predict *both* x and y. The first regression equation, $y=r\cdot\frac{\sigma_y}{\sigma_x}\cdot x$, is to be used *only* when y is to be predicted from x (when y is the "dependent" variable), while the second regression equation, $x=r\cdot\frac{\sigma_x}{\sigma_y}\cdot y$, is to be used *only* when x is to be predicted from y (when x is the "dependent" variable).[1] There are always two regression equations unless the correlation is perfect. When $r=1.00$, however, the equation $y=r\cdot\frac{\sigma_y}{\sigma_x}\cdot x$ becomes $y=\frac{\sigma_y}{\sigma_x}.x$, or $\sigma_x\cdot y=\sigma_y\cdot x$, while the equation $x=r\cdot\frac{\sigma_x}{\sigma_y}\cdot y$ becomes $x=\frac{\sigma_x}{\sigma_y}\cdot y$, or $\sigma_x\cdot y=\sigma_y\cdot x$. The two equations are now identical, and the regression lines coincide.

As an illustration of this last condition suppose that the

[1] A dependent variable depends for its value on the other variable in the equation. Thus in the equation $y=r\cdot\frac{\sigma_y}{\sigma_x}\cdot x$, y "depends" on the value given x.

correlation between height and weight is perfect, σ_x and σ_y remaining the same. The first regression equation would now become $y=1.00\times\frac{6.55}{7.75}\cdot x$, or $y=.85y$, while the second regression equation would become $x=1.00\times\frac{7.75}{6.55}\cdot y$, or $x=1.18y$. Algebraically, $x=1.18y$ is equivalent to $y=.85x$ (since in the second equation $x=\frac{y}{.85}$, or $x=1.18y$). Under the prescribed conditions, therefore, we should have a single equation and a single line, which would represent equally well a change (deviation) in Y for a given change in X, or a change (deviation) in X for a given change in Y. It may be added that when $r=1.00$, and in addition the two σ's are equal or are made equal by the arrangement of the diagram, the single regression line makes an angle of 45 degrees with the horizontal axis (see Diagram XVIII, and the discussion on pages 161–162).

2. The Regression Equations in Score Form

In the last paragraph the point was stressed that formulas (28) and (29) are the equations of the regression lines in *deviation form*—that values of x and y substituted in these equations are deviations from the means of the X and Y distributions and not actual scores or measures.[1] While equations in *deviation form* are all that we actually need for purposes of prediction, it is often very convenient to be able to estimate an individual's actual score in Y, say, directly from his score in X without the trouble of first converting the X-score into a deviation from the mean X. This can be done very simply if we employ the *score form* rather than the *deviation form* of the regression equation. The conversion of *deviation to score form* may be made as follows. Let the average of the Y's be denoted by Y' and any Y-score by Y, then the y deviation of any individual from the mean will be $Y-Y'$ (the difference between

[1] The small letters x and y are used to denote deviations from the means of the X and Y distributions. The large letters X and Y denote actual scores.

the score and the mean) or, in general, $y=Y-Y'$. In the same way, we can show that, in general, $x=X-X'$, when x is the deviation of any X score from the mean X from X'.

Now substitute $Y-Y'$ for y and $X-X'$ for x in formulas (28) and (29) and the two regression equations become,

$$Y-Y'=r\cdot\frac{\sigma_y}{\sigma_x}(X-X') \text{ or } Y=r\cdot\frac{\sigma_y}{\sigma_x}(X-X')+Y', \quad . \quad (30)$$

and

$$X-X'=r\cdot\frac{\sigma_x}{\sigma_y}(Y-Y') \text{ or } X=r\cdot\frac{\sigma_x}{\sigma_y}(Y-Y')+X', \quad . \quad (31)$$

These are the equations of the two regression lines in *score form.* In both equations, X and Y now represent *actual scores* and not *deviations* from the means of the two distributions.

If we substitute in (30) the values for Y', r, σ_y, σ_x, and X' obtained from Diagram XXIII, the equation becomes

$$Y-172.6=.60\times\frac{6.55}{7.75}(X-63.4),$$

or, clearing of fractions,

$$Y=.51X+140.3.$$

To illustrate the use of this equation, let us suppose that a man in our group weighs 60 kgs. (X) and that we wish to estimate his most probable height (Y). Substituting 60 for X in the equation, $Y=170.9$; and accordingly the most probable height of a man who weighs 60 kgs. is 170.9 cms.

If the problem is to predict weight instead of height, we must use equation (31). Substituting the values for X', r, σ_y, σ_x, and Y' in the second equation we have

$$X-63.4=.60\times\frac{7.75}{6.55}(Y-172.6)$$

or

$$X=.71Y-59.1.$$

Now given a man 180 cms. tall, we find putting 180 for Y in the formula, that $X=68.7$ kgs. Hence the most probable weight of a man 180 cms. tall is 68.7 kgs.

It may seem strange to the student to talk of "predicting" a man's height from his weight, when we already know the height and weight of all 120 men in our group. Of course when we have both height and weight it is unnecessary to convert one into the other. Suppose, however, that all we know about a certain man is his weight and the fact that he falls within the age-range of our group of 120 men. Now since we know the correlation between height and weight in this group it is possible from the regression equation to predict the most probable height of our subject in lieu of actually measuring him. In the same way, the regression equation may be used to predict the height of any man in the population from which our group is taken, provided our group is a random sample of the larger group. The regression equations hold, of course, *only* for the population from which the sample group is drawn. We could not, of course, estimate the probable heights of children or of women from a regression equation which had been worked out for men between the ages of 18 and 25 (the age-range of the men in our group of 120). And conversely, we could not expect regression equations worked out for elementary children to hold for older groups.

Probably height and weight—since they are both easily measured—do not show the value of the regression equations as well as other and more complex traits. To take a problem of more direct interest, suppose that in a group of children of approximately the same age the r between IQ and average grades made in the first year of high school works out to be .70. Now if we know the IQ of a child entering school the next year, it is possible to estimate what his probable scholastic performance will be from the regression equation worked out from the group of the previous year. This may be extremely valuable in educational guidance. The same thing is true of vocational guidance—we may be able on the basis of test scores to predict the probable success of an individual who contemplates entering a certain trade or profession, and thus advise him more intelligently.

3. The Reliability of the Predictions Made by the Regression Equations

A. The Standard Error of Estimate, $\sigma_{(\text{est.})}$, or S

We have constantly referred to the values of X and Y "predicted" from the regression equations as being the "most probable" values of the one variable accompanying the given value of the other. The method of showing just how reliable, i.e., how probable, our predicted values are, is to calculate their standard error of estimate, written $\sigma_{(\text{est})}$. To find the accuracy with which we are able to estimate Y-values from equation (30), we employ the formula [1]

$$\sigma_{(\text{est. } Y)} = \sigma_y \sqrt{1 - r^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

in which σ_y is the σ of the Y-distribution, and the "(est.)" is to distinguish its σ from the expressions $\sigma_{(\text{dis.})}$, $\sigma_{(\text{aver.})}$, etc., r is, of course, the coefficient of correlation between X and Y.

Now from equation (30) we have found that a man weighing 60 kgs. is most probably 170.9 cms. tall (see page 181). To find the reliability of this estimate substitute in formula (32), to find,

$$\sigma_{(\text{est. } Y)} = 6.55 \times \sqrt{1 - .6^2} = 5.2.$$

We may now say that the most probable height of a man weighing 60 kgs. is 170.9 cms. with a $\sigma_{(\text{est.})}$ of 5.2 cms.—and that the chances are 68 in 100 that the actual height of the given individual falls within the limits 170.9 ± 5.2, or between 165.7 cms. and 176.1 cms. We may be practically certain that the height of this man falls within the limits $170.9 \pm 3 \times 5.2$; or between 155.3 cms. and 186.5 cms.

In order to find with what degree of accuracy we are able to predict X values from equation (31) we use the formula, [2]

$$\sigma_{(\text{est. } X)} = \sigma_x \sqrt{1 - r^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

in which σ_x is the σ of the X-distribution.

[1] $\sigma_{(\text{est. } Y)}$ is sometimes written S_y.
[2] $\sigma_{(\text{est. } X)}$ is sometimes written S_x.

We have already found from formula (31) that the most probable weight (X) of a man 180 cms. tall is 68.7 kgs. (see page 181). To find the $\sigma_{(\text{est. }X)}$ of this prediction we substitute for σ_x and r in formula (33):

$$\sigma_{(\text{est. }X)} = 7.75 \times \sqrt{1 - .6^2} = 6.2.$$

Hence the most probable weight of a man in our group (or in the population from which it is drawn) who is 180 cms. tall is 68.7 kgs. with a $\sigma_{(\text{est.})}$ of 6.2 kgs. The chances are 68 in 100 that the actual weight of this man falls within the limits 68.7 ± 6.2, or between 62.5 and 74.9 kgs. We may be practically certain that his weight falls within the limits $68.7 \pm 3 \times 6.2$ or between 50.1 and 87.3 kgs.

B. The Probable Error of Estimate, $PE_{(\text{est.})}$

The $PE_{(\text{est.})}$ may be used for estimating the accuracy of a prediction instead of $\sigma_{(\text{est.})}$. $PE_{(\text{est.})}$ is obtained by simply multiplying $\sigma_{(\text{est.})}$ by the constant .6745. Thus

$$PE_{(\text{est. }Y)} = .6745 \times \sigma_y \sqrt{1 - r^2}, \quad . \quad . \quad . \quad . \quad (34)$$

and

$$PE_{(\text{est. }X)} = .6745 \times \sigma_x \sqrt{1 - r^2}, \quad . \quad . \quad . \quad . \quad (35)$$

The height of a man who weighs 60 kgs. has been estimated to be 170.9 cms. with a $\sigma_{(\text{est. }Y)}$ of 5.2 cms. The $PE_{(\text{est. }Y)}$ of this estimated height is $.6745 \times 5.2$ or 3.5 cms. The chances are even, therefore, that the actual height of this man falls within the limits 170.9 ± 3.5 or between 167.4 and 174.4 cms.

In like manner, since the estimated weight of a man 180 cms. tall is 68.7 kgs. with a $\sigma_{(\text{est. }X)}$ of 6.2, the $PE_{(\text{est. }X)}$ of this man's weight will be $.6745 \times 6.2$ or 4.2 kgs. The chances are even that this man's actual weight lies within the limits 68.7 ± 4.2 or between 64.5 and 72.9 kgs.

The formulas for $\sigma_{(\text{est.})}$ and $PE_{(\text{est.})}$ measure the error made in taking predicted instead of actual X and Y scores. Note that when $r = 1.00$, $\sqrt{1 - r^2}$ is 0; and consequently since both

$\sigma_{(est.)}$ and $PE_{(est.)}$ are then zero, there is no error of prediction. This result follows because all of the paired scores fall on the one double regression line when $r=1.00$[1] (see page 161).

An inspection of the formulas for $\sigma_{(est.)}$ and $PE_{(est.)}$ shows that the accuracy of the prediction from the regression equations depends upon the σ's of the two distributions (the σ_y and σ_x) and upon the degree of correlation between the two traits. If the variability in Y, say, is small, and the correlation between Y and X high (e.g., .90 to 1.00) values of Y can be predicted from known values of X with a comparatively high degree of accuracy. When the variability is large or the correlation low, however, the prediction often becomes so unreliable as to be almost valueless; and even with a fairly high coefficient, predictions will often have such a large error of estimate as to be almost valueless. Thus, in spite of the fact that an $r=.60$ is usually considered fairly substantial,[2] we can only predict a man's height (Y), knowing his weight (X), within a $PE_{(est.)}$ of 3.5 cms. In other words, the chances are only 50 in 100 that the actual height does not differ from the predicted height by more than ± 3.5 cms.

When using the regression equations for prediction, the $\sigma_{est.}$ or the $PE_{est.}$ should always be given. In general, the value of a prediction will depend—in addition to the size of the error of estimate—upon the fineness of the units of measurement and the purposes for which the prediction is made.

VI. The Complete Solution of a Correlation Problem

In Diagram XXIV will be found the complete solution of a second correlation problem. The purpose of another "model" problem, in addition to the height-weight problem in Diagram XXIII, is to strengthen the student's grasp on correlation by having him work through the steps in finding r and the regression equations with a new set of data. Often-

[1] See Monroe, *An Introduction to the Theory of Educational Measurements*, 1923, pp. 351–353, for a graphical demonstration of the meaning of $\sigma_{(est.)}$.

[2] See, however, the discussion of high and low correlation on page 288ff.

DIAGRAM XXIV

To Illustrate the Complete Solution of a Correlation Problem

IQ First Test (*X*-variable)

IQ Second Test (*Y*-variable)	90-94	95-99	100-104	105-109	110-114	115-119	120-124	125-129	130-134	135-139	140-144	145-149	150-154	Fy	Dy	FDy	FD_y^2	$\Sigma x'$ +	$\Sigma x'$ −	$\Sigma x'y'$ +	$\Sigma x'y'$ −
155-159											1	1	1	3	8	24	192	13		144	
150-154												1	1	2	7	14	98	13		91	
145-149												1	1	2	6	12	72	13		78	
140-144										4	3	1		8	5	40	200	37		185	
135-139									1	4	1			6	4	24	96	24		96	
130-134							1	1	3	1	1			7	3	21	63	21		63	
125-129					1	1	1	7	1	1	1			13	2	26	52	26		52	
120-124					3	2	3	3	1	1				13	1	13 (174)	13	13		13	
115-119					3	16	5		2					26	0						
110-114				2	7	5	2	3						19	−1	−19	19		3	3	
105-109			2	3	3	2	2							12	−2	−24	48		13	26	
100-104			4	8	3									15	−3	−45	135		31	93	
95-99		2	2	1	1									6	−4	−24	96		17	68	
90-94	2	1												3	−5	−15	75		14	70	
85-89	1													1	−6	−6(−133)	36		5	30	
																41	1195			1012	
Fx	3	3	8	14	21	26	14	14	8	11	7	4	3	136							
Dx	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7								
FDx	−15	−12	−24	−28	−21	(−100)	14	28	24	44	35	24	21(190)	=90							
FD_x^2	75	48	72	56	21		14	56	72	176	175	144	147	=1056							

$$cy=\frac{41}{136}=.3 \qquad c^2y=.09 \qquad Cy=1.5$$
$$My=117.5+1.5=119$$
$$\sigma y=\sqrt{\frac{1195}{136}-.09}\times 5=2.95\times 5=14.75$$

$$cx=\frac{90}{136}=.66 \qquad c^2x=.44 \qquad Cx=3.30$$
$$Mx=117.5+3.30=120.8$$
$$\sigma x=\sqrt{\frac{1056}{136}-.44}\times 5=2.71\times 5=13.55$$

Calculation of r:

$$r=\frac{\frac{1012}{136}-.3\times 66}{2.95\times 2.71}=.91$$

$PE_r=.01$ (Table XVIII)

Calculation of Regression Equations:

I. Deviation Form:

$$y=.91\times\frac{14.75}{13.55}x=.99x$$
$$x=.91\times\frac{13.55}{14.75}y=.84y$$

II. Score Form:

$$Y-119=.99(X-120.8)$$
$$Y=.99X-.59$$
$$X-120.8=.84(Y-119)$$
$$X=.84Y+20.8$$

Calculation of $PE_{(\text{est.})}$

$$PE_{(\text{est. } Y)}=.6745\times 14.75\times\sqrt{1-(.91)^2}=4.12(4)$$
$$PE_{(\text{est. } X)}=.6745\times 13.55\times\sqrt{1-(.91^2)}=3.79(4)$$

Examples:

Let $X=100$
$Y=99-.59$ or 98 ± 4

Let $X=120$
$Y=118\pm 4$

Let $Y=100$
$X=84+20.84$
$=104\pm 4$

times when only a single model problem is given, one fails to understand certain points in the solution which another entirely different problem will succeed in clearing up. A brief discussion of the important points in the solution of this problem will be given in the following paragraphs, which the student should read with Diagram XXIV before him.

The problem is to find the relation between the *IQ*'s of 136 children (of same chronological age) as determined from two individual intelligence tests. The correlation table has been constructed from a scatter diagram as explained on page 154. The first set of *IQ*'s is the X-variable, and the second set of *IQ*'s the Y-variable. Since the calculations of the two averages, c_x, c_y, σ_x, and σ_y, cover familiar ground and have been given in detail on the diagram, they need not be repeated.

Note first, then, that the product-deviations in the $\Sigma x'y'$ column have been taken from column 115–119 (the column containing the *GA* of the X-distribution) and row 115–119 (the row containing the *GA* of the Y-distribution). The entries in the $\Sigma x'y'$ column have been obtained by the shorter method described on page 167—each cell frequency in a given row has been multiplied by its D_x, and the sum of these partial deviations entered in the column $\Sigma x'$. This entry has then been "weighted" (multiplied) once for all by the D_y of the whole row. To illustrate, in the first row (reading from left to right) we have $(1\times5)+(1\times6)+(1\times7)$, or 18, as $\Sigma x'$ entry. (The D_x's are 5, 6, and 7, respectively, and may be found from the D_x row at the bottom of the diagram.) The common D_y is 8, hence the $\Sigma x'y'$ entry is 18×8 or 144. Again in the eighth row, we have $(3\times-1)+(2\times0)+(3\times1)+(3\times2)+(1\times3)+(1\times4)$ or 13 as the $\Sigma x'$ entry. The D_y of this row is 1, and hence the $\Sigma x'y'$ entry is 13. To take still another example, in the eleventh row we have $(2\times-3)+(3\times-2)+(3\times-1)+(2\times0)+(2\times1)$ or -13 as the $\Sigma x'$. Since the common D_y is (-2), the $x'y'$ entry here is $+26$.

After all of the $\Sigma x'y'$ entries have been made and the sum of the column found, the calculation of r from formula (23) and of

PE_r from formula (26) are simply matters of substitution. Remember that c_x, c_y, σ_y, σ_x, are all left in *units of step-interval* in the r formula (see page 167).

The regression equations in Deviation Form under (1) have been found by substituting the values of r, σ_x, and σ_y in formulas (28) and (29), and the two straight lines which these two equations represent have been plotted on the diagram. So far as the actual solution of the problem is concerned, it is unnecessary to plot these lines. They are of value, however, in indicating whether the means of the X and Y arrays may be fairly represented by straight lines; i.e., whether the regression is apparently "linear." If the relation is not "straight-line," other methods must be employed in calculating the correlation (see page 203.)

The regression equations in Score Form have been found, the one by substituting the two averages and the regression coefficient of Y on X (.99) in formula (30), and the other by substituting the two averages and the regression equation of X on Y (.84) in formula (31). The calculation of the two PE's of estimate is shown on the Diagram. $PE_{(\text{est. }Y)}$ is found from formula (34); $PE_{(\text{est. }X)}$ from formula (35).

Several examples have been given in the diagram to illustrate the use of the regression equations in "prediction." Note that an IQ of 100 on the first test (X) is most probably accompanied by an IQ of 98 on the second test (Y) with a $PE_{(\text{est. }Y)}$ of 4.12 (4) points. The chances are 50 in 100 that the actual IQ on the second test falls within the limits 98 ± 4, or between 102 and 94. An IQ of 120 on the first test (X) is most probably accompanied by an IQ of 118 points in the second test (Y), and the $PE_{(\text{est. }Y)}$ is again 4 points. All predicted Y's have the same error of estimate, no matter where on the scale the Y may fall.

While the errors of estimate $\sigma_{(\text{est.})}$ and $PE_{(\text{est.})}$ have been used hitherto for the purpose of giving the reliability of *specific* predicted scores, they may also be interpreted in a more general fashion. A $PE_{(\text{est. }Y)}$, for instance, of 4 points may be

taken to mean that one half of the IQ's in test Y failed of perfect correlation with the IQ's in test X by ± 4 points or more, while the other one half failed of perfect correlation by less than ± 4 points.

In most correlation problems we are interested in predicting the scores on only one test. (Y is usually taken as the dependent, and X the independent variable.) For illustrative purposes, however, an example is given in Diagram XXIV of the prediction of an IQ in X from an IQ in Y. Thus for an $IQ(Y)$ of 100 we find the most probable $IQ(X)$ to be 104 with a $PE_{(\text{est. } X)}$ of 3.79 (4) points. The chances are 50 in 100 that the actual $IQ(X)$ falls within the limits 104 ± 4 points or between 100 and 108.

VII. Methods of Measuring Correlation Which Take Account Only of Relative Position or Rank

In many problems, especially in the fields of applied and vocational psychology, the investigator finds that he must work with data in which differences in capacity or merit are expressed in *ranks* rather than in graded scores or measures. To mention a few cases of this sort, we have individuals ranked in order of merit for honesty, athletic ability, salesmanship, or intelligence; and advertisements, colors, etc., ranked for esthetic qualities, beauty, or individual preference. In computing correlations from such material as this it is neccessary to use methods which take account only of the relative positions or ranks. Also, when we have only a few scores (10 to 25 for example), it is often advisable to rank these in orders of merit and compute the correlation by a rank method instead of by the longer and more laborious product-moment method. Coefficients of correlation calculated from a few cases are nearly always unreliable, and of little value except in suggesting the possible existence of relation, or as a preliminary survey. In such cases, therefore, simple methods are recommended, as they save much time and labor besides giving

results which are as good as those secured by more elaborate methods.

In the present Section we shall consider two methods of finding the correlation when the data to be correlated have been arranged in orders of merit. These methods are known respectively as (1) the Method of Rank-Differences, and (2) the Method of Gains or the Spearman "Footrule."

1. The Method of Rank-Differences

The method of rank-differences is illustrated in Table XIX. The problem is to find the relation between the length of service and the selling efficiency of 12 salesmen. The men are listed in column 1, and in column 2, opposite the name of each man, is given the number of years he has been in the service of the company. In column 3, the men are ranked in order of merit in accordance with the length of their service. For example, G who has been longest with the company is ranked 1; C, the next longest, is ranked 2; and so on down the list. Notice that both A and J have the same period of service, and that each is ranked 7.5. Instead of ranking one 7, and the other 8, or both 7 or 8, we compromise by ranking both 7.5, and F who follows 9.[1]

In column 4 the men are ranked in order of merit for efficiency by the salesmanager. The most efficient man (C) is ranked 1, the least efficient (B) is ranked 12. In column 5, the difference (the "D") between each man's efficiency rank and his years of service rank is entered, and in the next column (6) each of these D's is squared. The correlation between the two orders of merit may now be computed by substituting for ΣD^2 and N in the formula,

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}, \quad \ldots\ldots \quad (36)$$

[1] When three or more individuals (or specimens of any sort) are tied—have the same score—the simplest plan is to give them all the median order of merit rating. Thus three individuals who are 5, 6, and 7, respectively, are all ranked 6, and the next following 8; while four individuals who are 5, 6, 7, and 8, are all ranked 6.5, and the next following 9.

TABLE XIX

To Illustrate the Rank-Difference Method of Finding Correlation

(1) Salesmen	(2) Years of Service	(3) Order of Merit (Service)	(4) Order of Merit (Efficiency)	(5) Difference between Ranks (D)	(6) Difference Squared (D^2)
A	5	7.5	6	1.5	2.25
B	2	11.5	12	.5	.25
C	10	2	1	1.0	1.00
D	8	4	9	5.0	25.00
E	6	6	8	2.0	4.00
F	4	9	5	4.0	16.00
G	12	1	2	1.0	1.00
H	2	11.5	10	1.5	2.25
I	7	5	3	2.0	4.00
J	5	7.5	7	.5	.25
K	9	3	4	1.0	1.00
L	3	10	11	1.0	1.00
$N=12$					58.00

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)} = 1 - \frac{6\times 58}{12(143)} = .80$$

From Table XX $r = .81$.

$$PE_r = \frac{.7063(1-r^2)}{\sqrt{N}} = .07 \qquad \text{[See formula (37)]}$$

in which D represents the difference in the rank of an individual in the two series, and ΣD^2 is the sum of the squares of all such differences. N is, of course, the number of cases, and ρ is the rank order coefficient of correlation. ρ may be transmuted into a product-moment r by means of Table XX.

Substituting 58 for ΣD^2 and 12 for N in formula (36), we obtain a ρ of .80, and from Table XX this is found to be equivalent to an r of .81. The PE of an r found from a ρ, is about 5% larger than the PE of the product-moment r.[1] The formula is

$$PE_r = \frac{.7063(1-r^2)}{\sqrt{N}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

and since, in the present example, $r = .81$, $PE_r = .07$. Accordingly, the coefficient of correlation though based on only 12

[1] See Brown & Thomson, *Essentials of Mental Measurement*, 1921, p. 103.

cases is conventionally reliable. Whenever N is less than 30, however, the PE_r is probably much larger than the value given by the formula. In any case r's and PE_r's secured from less than 30 cases should be accepted as tentative, and interpreted with caution. In the present example, all that we are justified in concluding is that in our particular group of 12 men there is evidence of a close correspondence between rankings for efficiency and number of years employed.

TABLE XX

A Table to Infer the Value of r from Any Given Value of ρ

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

ρ	r	ρ	r	ρ	r	ρ	r
.01	.0105	.26	.2714	.51	.5277	.76	.7750
.02	.0209	.27	.2818	.52	.5378	.77	.7847
.03	.0314	.28	.2922	.53	.5479	.78	.7943
.04	.0419	.29	.3025	.54	.5580	.79	.8039
.05	.0524	.30	.3129	.55	.5680	.80	.8135
.06	.0628	.31	.3232	.56	.5781	.81	.8230
.07	.0733	.32	.3335	.57	.5881	.82	.8325
.08	.0838	.33	.3439	.58	.5981	.83	.8421
.09	.0942	.34	.3542	.59	.6081	.84	.8516
.10	.1047	.35	.3645	.60	.6180	.85	.8610
.11	.1151	.36	.3748	.61	.6280	.86	.8705
.12	.1256	.37	.3850	.62	.6379	.87	.8799
.13	.1360	.38	.3935	.63	.6478	.88	.8893
.14	.1465	.39	.4056	.64	.6577	.89	.8986
.15	.1569	.40	.4158	.65	.6676	.90	.9080
.16	.1674	.41	.4261	.66	.6775	.91	.9173
.17	.1778	.42	.4363	.67	.6873	.92	.9269
.18	.1882	.43	.4465	.68	.6971	.93	.9359
.19	.1986	.44	.4567	.69	.7069	.94	.9451
.20	.2091	.45	.4669	.70	.7167	.95	.9543
.21	.2195	.46	.4771	.71	.7265	.96	.9635
.22	.2299	.47	.4872	.72	.7363	.97	.9727
.23	.2403	.48	.4973	.73	.7460	.98	.9818
.24	.2507	.49	.5075	.74	.7557	.99	.9909
.25	.2611	.50	.5176	.75	.7654	1.00	1.0000

2. The Method of Gains, or the Spearman Footrule

A second method of computing correlation when the data are ranked in orders of merit is the Method of Gains, or the Spearman "Footrule." Table XXI illustrates the use of the Foot-

rule with the data taken from Table XIX. It will be noticed that the first four columns are the same in both methods, i.e., each series is arranged first in an order of merit. The methods differ from here on, however. The entries in column 5, which is headed G ("Gains"), are found by taking the *plus* differences or the *gains* in rank of the 12 men in the efficiency-rankings as compared with their service-rankings. Thus A who ranks 7.5 in "service" and 6 in "efficiency" has an increase in rank or gain of 1.5 in the second ranking over the first.[1] C, F, H, I, and J, likewise register plus differences or gains in their efficiency rankings as compared with their service rankings. The total of the G column is 10.5. Note that if we compute the gains in rank of service over efficiency instead of efficiency over service, the same G will be obtained. This is shown in column 6, marked G'. It makes no difference, therefore, whether we figure gains of the first series over the second, or the other way round, second over first.

TABLE XXI

To Illustrate the Footrule Method of Finding Correlation

(1) Salesmen	(2) Years of Service	(3) Order of Merit (Service)	(4) Order of Merit (Efficiency)	(5) G (Gains) (4 over 3)	(6) G' (Gains) (3 over 4)
A	5	7.5	6	1.5	
B	2	11.5	12	....	.5
C	10	2	1	1.0	
D	8	4	9	....	5.0
E	6	6	8	....	2.0
F	4	9	5	4.0	
G	12	1	2	....	1.0
H	2	11.5	10	1.5	
I	7	5	3	2.0	
J	5	7.5	7	.5	
K	9	3	4	....	1.0
L	3	10	11	....	1.0
				10.5	10.5

$$R = 1 - \frac{6\Sigma G}{N^2 - 1} = 1 - \frac{6 \times 10.5}{143} = .56$$

r (Table XXII) $= .79$

[1] Since the rankings are from 1 to 12, a rank of 6 is to be taken as higher than a rank of 7.5.

When the sum of the G column has been obtained, the correlation may be found from the formula,

$$R = 1 - \frac{6\Sigma G}{(N^2 - 1)}, \quad \ldots \ldots \quad (38)$$

Substituting for ΣG its value 10.5, and for N its value 12, we get an R of .56. From Table XXII this R may be converted into an equivalent product-moment r of .79. Note that this value of r compares favorably with the r (found from ρ) of .81.

TABLE XXII

A TABLE TO INFER THE VALUE OF r FROM ANY GIVEN VALUE OF R

R	r	R	r	R	r	R	r
.00	.000						
.01	.018	.26	.429	.51	.742	.76	.937
.02	.036	.27	.444	.52	.753	.77	.942
.03	.054	.28	.458	.53	.763	.78	.947
.04	.071	.29	.472	.54	.772	.79	.952
.05	.089	.30	.486	.55	.782	.80	.956
.06	.107	.31	.500	.56	.791	.81	.961
.07	.124	.32	.514	.57	.801	.82	.965
.08	.141	.33	.528	.58	.810	.83	.968
.09	.158	.34	.541	.59	.818	.84	.972
.10	.176	.35	.554	.60	.827	.85	.975
.11	.192	.36	.567	.61	.836	.86	.979
.12	.209	.37	.580	.62	.844	.87	.981
.13	.226	.38	.593	.63	.852	.88	.984
.14	.242	.39	.606	.64	.860	.89	.987
.15	.259	.40	.618	.65	.867	.90	.989
.16	.275	.41	.630	.66	.875	.91	.991
.17	.291	.42	.642	.67	.882	.92	.993
.18	.307	.43	.654	.68	.889	.93	.995
.19	.323	.44	.666	.69	.896	.94	.996
.20	.338	.45	.677	.70	.902	.95	.997
.21	.354	.46	.689	.71	.908	.96	.998
.22	.369	.47	.700	.72	.915	.97	.999
.23	.384	.48	.711	.73	.921	.98	.9996
.24	.399	.49	.721	.74	.926	.99	.9999
.25	.414	.50	.732	.75	.932	1.00	1.0000

The Footrule formula gives a rough estimate of the correlation, and is generally less accurate than the rank-difference formula. The coefficient R "has a large, though

except in the case of zero correlation, not definitely known PE; does not vary between -1 and $+1$; is not comparable in meaning with the product-moment coefficient; and in general has none of the merits except brevity of the formula based on the squares of the differences in rank." [1] The Footrule can be employed to advantage, however, when the data are so meager or crude as to make a more refined method a waste of time; or it may be used in a preliminary survey to determine whether there is sufficient evidence of correlation to warrant the application of the product-moment method.

3. Summary of the Rank Methods

The product-moment method takes account of both the size of the score and its position in the series. The rank methods take account only of the *position* of the items in the series. For example, individuals who score 90, 86, and 70, on a given test must be ranked 1, 2, and 3 in order of merit despite the fact that the difference between 90 and 86 is 4, and the difference between 86 and 70 is 16. The rank methods indicate the presence of relationship rather than the extent of relation. In general it may be set down as a convenient rule that rank methods should never be used ordinarily except when N is small—say less than 30. Of the two rank methods, the method of rank-differences is to be preferred as the more accurate.

VIII. A Method of Measuring Relationship When the Data are Grouped into Classes or Categories. The Contingency Method

Sometimes the need arises of computing correlation when the facts in which we are interested cannot be conveniently measured, but can be grouped into classes or categories. To cite a few examples of such data, we can classify eye-color as blue, grey, or brown; temper as quick, even, or slow; athletic

[1] See Kelley, T. L., *Statistical Method*, 1923, p. 193.

ability as good, average or poor, when we are unable to measure such facts exactly. The methods of computing correlation which have been given in the preceding sections are generally applied to facts which can be measured absolutely in terms of some common unit, or which, at least, can be ranked in order of merit—they do not ordinarily apply to data which can only be grouped into classes. Several methods are available for such material, however. One of the best of these is the Contingency Method developed by Prof. Karl Pearson.[1] In the contingency method relation is expressed by C, the Coefficient of Mean Square Contingency.

Table XXIII illustrates the method of drawing up a contingency table, and shows in detail the steps involved in finding C. The problem is to discover whether there is any "resemblance" (correlation) between the eye-color of father and son. There are 1000 cases. Tabulation of data is similar to the method used in constructing a correlation table. Reading down the first column, for example, we find that out of a total of 358 blue-eyed fathers, 194 have blue-eyed sons; 83 grey-eyed sons; 25 dark grey or hazel-eyed sons; and 56 brown-eyed sons. In the first row, we find 335 blue-eyed sons of whom 194 have blue-eyed fathers; 70 grey-eyed fathers; 41 dark grey or hazel-eyed fathers; 30 brown-eyed fathers.

After the contingency table is completed, the first step in the calculation of C is to find an "independence value" for each cell. These values—the figures in the parentheses in the cells—represent the number of fathers and sons (whose eye-color is given by the column and row, respectively, in which the cell lies) whom we should expect to find in any given cell in the absence of any *actual association* in the eye-color of father and son. For example, the observed number of blue-eyed fathers who have blue-eyed sons in our sample of 1000 is 194. If there were no correlation between the eye-color of father and son, we should still expect to find $\frac{335 \times 358}{1000}$ or

[1] Yule, G. U., *An Introduction to the Theory of Statistics*, 1919, p. 64ff.

TABLE XXIII

To Illustrate the Calculation of C, the Coefficient of Mean Square Contingency. [From Yule, p. 70]

Father's Eye Color

Son's Eye Color	Blue	Grey	Hazel	Brown	Totals
Blue	(120) 194	(88) 70	(60) 41	(66) 30	335
Grey	(102) 83	(75) 124	(51) 41	(56) 36	284
Hazel	(49) 25	(36) 34	(25) 55	(27) 23	137
Brown	(87) 56	(64) 36	(44) 43	(48) 109	244
Totals	358	264	180	198	1000

Column 1

Independence Values

$$\frac{335\times358}{1000}=120 \qquad \frac{137\times264}{1000}=36$$

$$\frac{335\times264}{1000}=88 \qquad \frac{137\times180}{1000}=25$$

$$\frac{335\times180}{1000}=60 \qquad \frac{137\times198}{1000}=27$$

$$\frac{335\times198}{1000}=66 \qquad \frac{244\times358}{1000}=87$$

$$\frac{284\times358}{1000}=102 \qquad \frac{244\times264}{1000}=64$$

$$\frac{284\times264}{1000}=75 \qquad \frac{244\times180}{1000}=44$$

$$\frac{284\times180}{1000}=51 \qquad \frac{244\times198}{1000}=48$$

$$\frac{284\times198}{1000}=56$$

$$\frac{137\times358}{1000}=49$$

Column 2

$$\frac{(194)^2}{120}=313.6$$
$$\frac{(83)^2}{102}=67.5$$
$$\frac{(25)^2}{49}=12.8$$
$$\frac{(56)^2}{87}=36.0$$
$$\frac{(70)^2}{88}=55.7$$
$$\frac{(124)^2}{75}=205.0$$
$$\frac{(34)^2}{36}=32.1$$
$$\frac{(36)^2}{64}=20.3$$
$$\frac{(41)^2}{60}=28.0$$
$$\frac{(41)^2}{51}=33.0$$
$$\frac{(55)^2}{25}=121.0$$
$$\frac{(43)^2}{44}=42.0$$
$$\frac{(30)^2}{66}=13.6$$
$$\frac{(36)^2}{56}=23.1$$
$$\frac{(23)^2}{27}=19.6$$
$$\frac{(109)^2}{48}=247.5$$

$$S=1270.8$$
$$N=1000$$
$$S-N=270.8$$

$$C=\sqrt{\frac{S-N}{S}}=\sqrt{\frac{270.8}{1270.8}}=462$$

120 blue-eyed fathers with blue-eyed sons by the operation of chance alone.[1] Again, the observed number of grey-eyed fathers who have blue-eyed sons is 70. In the absence of any real association, chance alone would account for $\frac{335 \times 264}{1000}$ or 88 such cases in our sample of 1000. In like manner "independence values" may be found for each cell by the simple process of multiplying together the totals of the row and column in which the cell lies and dividing this product by N, the number of cases. (See column 1, Table XXIII.)

When the independence values have been calculated for each cell, the next step is to square each cell entry and divide this result by the independence value of that cell (see column 2). All quotients so found are totaled to give S (1270.8), and N(1000) is subtracted to give $S-N$. The coefficient of mean square contingency, C, may then be found from the formula,

$$C=\sqrt{\frac{S-N}{S}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

In the present problem, $C = .462$.

The steps in the computation of C may be summarized as follows:

1. Construct a contingency table as shown in Table XXIII.
2. Determine the "independence value" for each cell by multiplying together the totals of the row and column in which the cell falls and dividing this product by N.
3. Square the number found in each cell, and divide this result by the independence value of that cell obtained in (2) above.
4. Sum the quotients obtained from (3). Call this total S.

[1] We find that $\frac{335}{1000}$ of all the sons are blue-eyed. This proportion should hold for sons of *all* fathers, if there is no dependence of son on father in respect to eye-color. Hence $\frac{335}{1000}$ of the 358 blue-eyed fathers should have blue-eyed sons by the operation of chance alone. This argument applies to the other "independence values" also.

5. Subtract N from S, giving $S-N$.

6. Divide $S-N$ by S and extract the square root to get C, the coefficient of mean square contingency.

The fundamental principle underlying the Contingency Method is a comparison of the frequency of association (number of cases) *actually* found in each cell with the frequency of association which we should *expect* to find in the cells if the traits considered were completely unrelated (independent). If there is just no correlation between the two variables in our contingency table, $C = .00$; if there is perfect correlation, C approaches 1.00 as a limit.

While in general no sign is attached to C, as this coefficient simply indicates whether the two traits are associated or independent, for interpretative purposes a minus sign may be affixed to a C if an inspection of the contingency table shows that marked degrees of the one trait are found with slight degrees of the other. Thus from an inspection of Table XXIII, it is evident that slight pigmentation of eyes in the father is associated with slight pigmentation of eyes in the son, and hence in the present case, C is clearly positive.[1] If marked pigmentation in the eyes of the father had been associated with slight pigmentation in the eyes of the son, C would have been negative. In other words, we must determine whether the correlation is positive or negative from the contingency table,—C gives simply the *degree* of the relation.

One disadvantage of the contingency method lies in the fact that C does not remain constant—for the same data—when the number of classes in the table is increased. The C calculated from a 3×3 fold table will not ordinarily equal the C calculated from the same data arranged in, say, a 5×5 fold table. Moreover, the maximum value which a C can take will depend

[1] Note, for example, that 194 blue-eyed fathers have blue-eyed sons, while only 30 brown-eyed fathers have blue-eyed sons. Also, 109 brown-eyed fathers have brown-eyed sons while only 56 blue-eyed fathers have brown-eyed sons. Other comparisons like these will show that association between the degree of pigmentation in the eyes of father and son is positive.

on the fineness of the classification employed. Yule[1] has shown that

when the number of classes = 2	C cannot exceed .707
when the number of classes = 3	C cannot exceed .816
when the number of classes = 4	C cannot exceed .866
when the number of classes = 5	C cannot exceed .894
when the number of classes = 6	C cannot exceed .913
when the number of classes = 7	C cannot exceed .926
when the number of classes = 8	C cannot exceed .935
when the number of classes = 9	C cannot exceed .943
when the number of classes = 10	C cannot exceed .949

Yule has suggested, in the light of these facts, that we "restrict the use of the 'coefficient of contingency' to 5×5-fold or finer classifications" in order that the maximum value of C may be as near unity as possible. On the other hand, we must avoid a too-fine classification or C will be affected by slight or "casual irregularities of no physical significance"; and in addition the arithmetic will be needlessly increased.

Since the classification in Table XXIII is 4×4-fold, the value of C would very probably change somewhat if the number of classes were increased. The table will serve very well, however, as an illustration of the method, and of the arithmetic involved in finding C. Moreover, as the maximum C from a 4×4-fold table is .866, and the C found from Table XXIII is .462, we are justified in concluding—in spite of the relative crudeness of our measures—that there is a medium positive correlation between pigmentation of eyes in father and son.

The relation of C to r, the Product-Moment coefficient of correlation, is of considerable importance. C may be taken as practically equivalent to r, (1) when the grouping is relatively fine,—5×5-fold or finer; (2) when the sample is large; (3) when we know, or are justified in assuming, that the traits which we are correlating are normally distributed. In case the first of these conditions is not fulfilled, Pearson[2] has given a correction for "broad categories" which should be used with 4×4-fold and less fine classifications, if C is to be compared with

[1] *An Introduction to the Theory of Statistics*, 1919, p. 66.

[2] Pearson Karl, *On the Measurement of the Influences of "Broad Categories" on Correlation.* Biometrika, Vol. IX, 1913.

r. For 5×5 fold or finer classifications this correction is usually small, and unless a very accurate measure of correlation is desired it may be disregarded and C taken as roughly equal to r.

TABLE XXIV

To Illustrate the Calculation of C by Short Method
Boys: Ages $4\frac{1}{2}$–$5\frac{1}{2}$ Years

Weight in Pounds

Height in Inches	24–28	29–33	34–38	39–43	44–48	49–53	Total
45–47			1		2		3
42–44			4	35	21	5	65
39–41		5	87	90	7	1	190
36–38	1	18	72	8			99
33–35	5	15	5				25
30–32	2						2
	8	38	169	133	30	6	384

Column 1: $\frac{1}{8}\left[\frac{1}{99}+\frac{25}{25}+\frac{4}{2}\right] = .3762$

Column 2: $\frac{1}{38}\left[\frac{25}{190}+\frac{324}{99}+\frac{225}{25}\right] = .3264$

Column 3: $\frac{1}{169}\left[\frac{1}{3}+\frac{16}{65}+\frac{7569}{190}+\frac{5184}{99}+\frac{25}{25}\right] = .5549$

Column 4: $\frac{1}{133}\left[\frac{1225}{65}+\frac{8100}{190}+\frac{64}{99}\right] = .4671$

Column 5: $\frac{1}{30}\left[\frac{4}{3}+\frac{441}{65}+\frac{49}{190}\right] = .2792$

Column 6: $\frac{1}{6}\left[\frac{25}{65}+\frac{1}{190}\right] = .0650$

$P = 2.0688$

$$C=\sqrt{\frac{P-1}{P}}=\sqrt{\frac{1.0688}{2.0688}}=.719$$

The arithmetic involved in computing C may be lessened somewhat by combining the twofold process of (1) calculating independence values and (2) dividing the square of each cell frequency by its independence value. This Short Method of finding C is illustrated in Table XXIV. Note that the first occupied cell in the *first* column of the table has a frequency of 1 and an independence value of $\frac{99\times 8}{384}$, and that the cell frequency squared and divided by the independence value is $\frac{1\times 384}{8\times 99}$. This quotient, viz., $\frac{1\times 384}{8\times 99}$ is the contribution of this particular cell to the total S. In like manner the contribution to S of the next cell in this column is $\frac{5^2\times 384}{8\times 25}$; and of the third and last cell, $\frac{2^2\times 384}{8\times 2}$. These contributions from column 1 may be combined as follows, $\frac{384}{8}\left(\frac{1}{99}+\frac{25}{25}+\frac{4}{2}\right)$; and the contribution of each of the other five columns to S may be found in exactly the same way. One further simplification may be made. Since $N(384)$ is a common factor in each column, it may be left out of the computations entirely in calculating the contribution of each cell, as shown in the table. Then if the sum of all six columns is denoted by P, $C=\sqrt{\frac{P-1}{P}}$ directly.[1]

By the Short Method, C is found to equal .719, and the coefficient of correlation for the same table will be found to be .709 (see page 216). The correspondence of C and r is somewhat closer here than is generally obtained, although the difference between C and r is never very great when the conditions prescribed on page 200 have been met. In the present

[1] Since $P=\frac{S}{N}$, $S=PN$. Substituting PN for S in the formula $C=\sqrt{\frac{S-N}{S}}$, $C=\sqrt{\frac{PN-N}{PN}}$ or removing the common factor, $C=\sqrt{\frac{P-1}{P}}$.

case, N is fairly large, the classification is 6×6-fold, and the distributions of both height and weight fairly normal.

The steps in the computation of C by the Short Method may be summarized as follows (see Table XXIV).

1. Square the frequency in each cell of column 1, and divide each square by the row total in which the cell falls.
2. Add all of the results for column 1, and divide by the column total, a common factor. Record this partial sum.
3. Repeat (1) and (2) for each of the other columns in the table.
4. Call the sum of all partial sums P.
5. Find C from the formula $C=\sqrt{\frac{P-1}{P}}$.

In many problems in psychology in which the relation between various attributes, whether of individuals or things, is sought, C will prove of considerable value.

IX. Non-Linear Relationship

1. The Correlation Ratio

The relation which exists between the paired values of two sets of measures X and Y may be described in a general way as either " linear " or " non-linear." When the means of the arrays of successive columns or rows in a correlation table follow straight lines (exactly or approximately) the regression is called " linear," and the relation between the two sets of measure or scores is a " straight line relation." On the other hand, when the drift or the trend of the means in the successive arrays cannot be described by a straight line, but can be properly represented only by a curve of some kind, the regression is called curvilinear, or in general non-linear, and the relation between the two variables is a " curved line relation."

Our previous discussion has been concerned entirely with cases in which the relation between X and Y was known to be linear and in which r gave a fair measure of the degree of correla-

tion. Cases sometimes arise in psychological measurement, however, in which the relation between X and Y is clearly non-linear, and in such cases the coefficient of correlation r—since the product-moment method assumes linear relationship—cannot be used. The reason for this may be stated in brief as follows. When a definitely curvilinear relation—instead of being described by a curve—is represented by a straight line, the scatter of the paired values is considerably greater about the straight line than about the curve. This results from the fact that the scatter about a curve joining the means of the successive arrays is necessarily less than the scatter about a straight line which has been "fitted" to these mean points. The less the scatter about the regression line or curve, the greater the degree of correlation; hence a coefficient of correlation calculated from a correlation table in which the regression is truly curvilinear will be materially less than the true correlation between the variables X and Y. (See Footnote 1.)

In order to measure non-linear relation, therefore, we need a more generalized coefficient than the coefficient of correlation, r:—that is, we need a coefficient which will measure the con-

[1] A simple illustration will make clear just why this is true. The correlation between the following two short series (Table XXV) by the product-moment formula (formula 25) is .93. The true correlation, however, is 1.00, i.e., perfect, since the Y values are absolutely dependent on the X values:—as X increases

TABLE XXV

Variable X	Variable Y
1	.25
2	.50
3	1.00
4	2.00
5	4.00

in steps of 1 (in arithmetic progression) Y doubles (increases in geometric progression). The reason why r is less than 1.00 is perfectly obvious as soon as we plot the paired X and Y values (see Diagram XXV). Since the relationship between X and Y is curvilinear, it cannot be described by a straight line. Consequently when straight line relationship is assumed (as in the product-moment formula) the plotted points do not fall on the relation line, and r is less than 1.00—the true correlation between X and Y. In true curvilinear correlation, r is always less than η.

centration of the paired X and Y values about a relation *curve*, just as r measures the concentration of the paired values about a relation *line*. One such coefficient is the Correlation Ratio, devised by Prof. Karl Pearson, and designated by the symbol η. (eta). Since eta is a general coefficient it may be employed when the regression is linear as well as non-linear. If the regression is linear—if the means of the arrays fall on straight lines —η will equal r; if the regression is non-linear—if the means

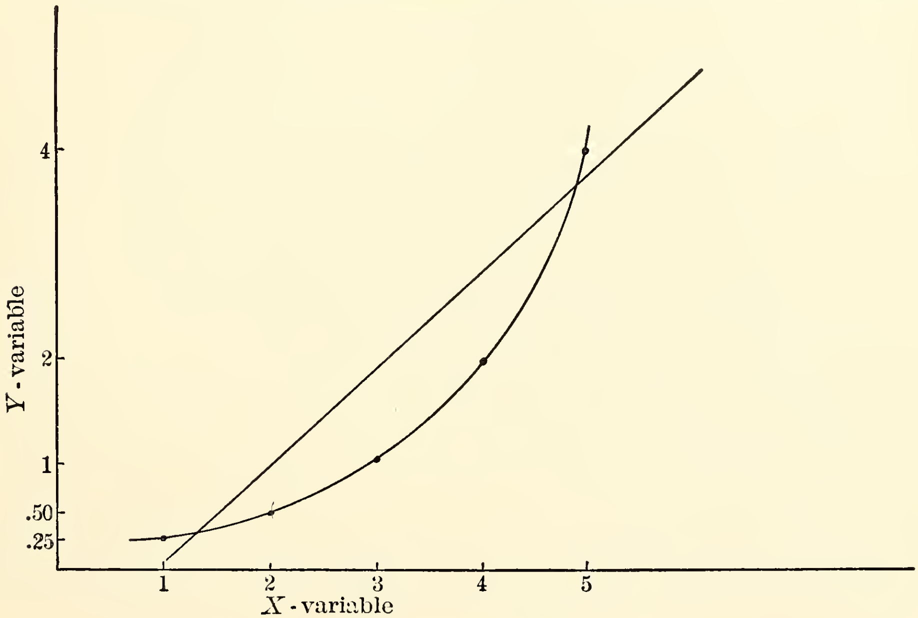


DIAGRAM XXV

do not fall on straight lines—η will be greater than r. In general, as long as the relation between Y and X is non-linear η and r will differ, η always being greater than r. The coefficient of correlation, therefore, is seen to be simply a limiting value of the more general η, just as straight line relationship is simply a limiting case of curvilinear relation.

η is always positive, and varies from zero to 1.00. Whether or not the *relation* given by η is positive, negative or a varying one must be determined, however, from the direction taken by the curve of relation; i.e., by inspection of the correlation diagram.

The process of calculating η from a correlation table in which the relation is definitely non-linear is shown in Diagram XXVI. The steps involved in finding the values to be substituted in the formula for η may be outlined as follows:

Step I

Construct a correlation table as shown in Diagrams XXIII and XXIV and described on page 154.

Step II

Find the average (Y') and the σ of the Y-distribution, using the Guessed average Method described in Chapter I.

Step III

Compute the averages (Y'_x) of the successive Y-arrays, i.e., the arrays of the columns. Enter these in row marked Y'_x.

Step IV

Find the deviation of each Y'_x from the average of the whole table, Y'; that is, find $(Y'_x - Y')$ for each column.

Step V

Square each deviation—each $(Y'_x - Y')$—and enter the results in the row marked $(Y'_x - Y')^2$.

Step VI

Multiply or weight each $(Y'_x - Y')^2$ by the F_x of its column. In the first column, for example, multiply 15.52 [i.e., $(Y'_x - Y')^2$] by 20, its F_x.

Step VII

Find the sum of the $F_x(Y'_x - Y')^2$ column. Divide this sum by N, and extract the square root. The result is σ_{my}, the standard deviation of the *means* of the various columns about the arithmetic mean of all of the Y's.

Step VIII

Divide σ_{my} by σ_y to get the correlation ration η_{yx}. The formula for η_{yx} may be written,

$$\eta_{yx} = \frac{\sigma_{my}}{\sigma_y}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

If now we substitute in formula (40) the values of σ_{my} and σ_y found from Diagram XXVI, the correlation-ratio η_{yx}

Number of problems worked Y-variable	Sp.	III	IV	V	VI	VII	VIII	IX	Fy	Dy	FDy	FD^2y
	Grade Position X-variable											
15								10	10	9	90	810
14								12	12	8	96	768
13								18	18	7	126	882
12							8	16	24	6	144	864
11							10	8	18	5	90	450
10							12		12	4	48	192
9							14		14	3	42	126
8							6		6	2	12	24
7						8	6		14	1	14 (667)	14
6						19	7		26	0		
5				2	2	22	5		31	−1	−31	31
4			1	10	17	26			54	−2	−108	216
3		2	4	8	15	12			41	−3	−123	369
2	5	5	12	8	24	9			63	−4	−252	1008
1	9	8	16	16	9	9			67	−5	−335	1675
0	6	6	3	20	13	7			55	−6	−330 (−1179)	1980
F_x	20	21	36	64	80	112	68	64	465		−517	9409
Y'_x	.95	1.14	1.56	1.66	2.25	4.00	9.00	13.00			$C_y=\frac{-517}{465}=-1.11$	$C_y^2=1.23$
(Y'_x-Y')	−3.94	−3.75	−3.33	−3.23	−2.64	−.89	4.11	8.11			$GA\,(Y)=6.00$	Average $=4.89\,(Y')$
$(Y'_x-Y')^2$	15.52	14.06	11.09	10.43	6.97	.79	16.89	65.77			$\sigma_y=\sqrt{\frac{9409}{465}-1.23}=4.36$	
$F_x(Y'_x-Y')^2$	310.40	295.26	399.24	667.52	557.60	88.48	1148.52	4209.28	= 7676.30			

$$\sigma_{my}=\sqrt{\frac{\Sigma[F_x(Y'_x-Y')^2]}{N}}=\sqrt{\frac{7676.30}{465}}=\sqrt{16.51}=4.06$$

$$\eta_{yx}=\frac{\sigma_{my}}{\sigma_y}=\frac{4.06}{4.36}=.931 \qquad PE_\eta=.004 \quad \text{(Table XVIII)}$$

DIAGRAM XXVI

To Illustrate Non-Linear Regression and the Calculation of the Correlation-Ratio, η_{yx}

becomes .931.[1] This coefficient shows how the number of problems worked (on the average) in a certain arithmetic test (Y) is related to the grade position (X) of 465 pupils. The curve which describes this relation—the curve which best marks the trend or "drift" of the means of the successive Y arrays—has been drawn in on the figure. Note that it begins low and gradually rises, suddenly bending up in a concave fashion.

From the diagram alone it would seem to be clear enough that the regression of Y on X is non-linear. Further evidence of this may be found in the fact that the coefficient of correlation, r, calculated from this table (on the assumption, of course, of linear relationship) is .80,—about .13 less than η_{yx}. The method of determining definitely whether regression is linear or non-linear in any table will be given in (**3**) following.

There are always two η's in every non-linear correlation table, just as there are always two regression coefficients, $r\frac{\sigma_y}{\sigma_x}$ and $r\frac{\sigma_x}{\sigma_y}$, in a table in which regression is linear. The one, written η_{yx}, refers to the regression of Y on X (Y is the dependent variable); the other, written η_{xy}, refers to the regression of X on Y (X is the dependent variable). The value of η_{xy} may be computed in exactly the same way as η_{yx} by substituting X for Y in the outline of "steps" given above. The formula is

$$\eta_{xy}=\frac{\sigma_{mx}}{\sigma_x}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

Unlike r which has the same value in both regression equations [see formulas (28) and (29)] η_{yx} and η_{xy} will usually differ, their values depending on the degree of scatter about the curves joining the means of the Y and X arrays. In the present

[1] The PE of η may be found from the formula

$$PE_\eta=\frac{.6745(1-\eta^2)}{\sqrt{N}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

or from Table XVIII.

problem, for example, $\eta_{xy} = .818$, while $\eta_{yx} = .931$ as shown above. In the special case in which the regression is truly linear, η_{yx} and η_{xy} equal each other, and both equal r (see page 205).

2. The Correction of "Raw" Eta

The value of η depends materially on the number of cases in the sample, and on the fineness of the grouping. As a general rule, η should never be calculated unless N is fairly large. When N is comparatively small or the number of arrays is large, Pearson[1] has given a correction which should be applied to the "raw" (i.e., calculated) value of η.

If we represent the number of arrays by κ the formula for "corrected eta" is

$$\text{corrected } \eta^2 = \frac{\eta^2 - \dfrac{(\kappa - 3)}{N}}{1 - \dfrac{(\kappa - 3)}{N}}, \quad \ldots \ldots \quad (43)$$

(The η on the *right* hand side of the equation is the "raw" eta.)

If we apply this correction to the value of η_{yx} obtained in the present problem, we have, substituting .931 for η_{yx}, 8 (the number of Y-arrays) for κ, and 465 for N,

$$\text{corrected } \eta^2_{yx} = \frac{(.931)^2 - .011}{1 - .011},$$

$$\bar{\eta}^2_{yx} = \frac{.856}{.989} = .8655,$$

and

$$\eta_{yx} = .930.$$

In the present case the correction is very small. If N is small, however, or κ large, the raw eta may be considerably reduced.

3. Test for Linearity of Regression

It is oftentimes difficult to tell from the appearance of a correlation table whether the regression is linear or non-linear,

[1] Biometrika, 1923, 14, 412–417.

and in such cases it is best to calculate both r and η. As stated above, if the regression is strictly linear η equals r; and the greater the departure from linearity the greater the difference between η and r. A simple test of linearity is that ζ (zeta) the difference between $\eta^2 - r^2$ shall differ from zero by an amount which is not greater than that which might arise from fluctuations due to random sampling. To make this test, we must first find PE_ζ given by the formula[1]

$$PE_\zeta = .6745 \times 2\sqrt{\frac{\zeta}{N}} \cdot \sqrt{(1-\eta^2)^2 - (1-r^2)^2 + 1}, \quad . \quad (44)$$

The second radical in formula (44) is approximately equal to 1, and hence unless great accuracy is required we may write the formula simply as

$$PE_\zeta = .6745 \times 2\sqrt{\frac{\zeta}{N}}, \quad . \quad . \quad . \quad . \quad . \quad (45)$$

In the problem which we have been considering $\eta_{yx} = .930$ and $r = .80$. Accordingly, $\zeta = (.930)^2 - (.80)^2$ or $.2249$, and from formula (45) $PE_\zeta = .030$.[2] Zeta, therefore, is 7.49 times its PE $\left(\text{since } \frac{\zeta}{PE_\zeta} = \frac{.2249}{.030} \text{ or } 7.49\right)$ and there is no doubt as to the non-linearity of the regression. To determine whether $\frac{\zeta}{PE_\zeta}$ denotes a real or simply a chance difference between η^2 and r^2, Table XV, the $\frac{D}{PE_{(\text{diff.})}}$ table, may be used conveniently.

If zeta is very small, or if both η and r are small, a simple test for linearity (Blakeman's test[3]) which does not require finding PE_ζ may be used. According to this test, when

$$N(\eta^2 - r^2) < 11.37 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

[1] This formula is due to Blakeman. See Yule, *An Introduction to the Theory of Statistics*, p. 352.

[2] Formula (44) gives PE (zeta) as .028. The difference between the results given by formulas (44) and (45) is negligible here.

[3] Blakeman, J., *On Tests for Linearity of Regression*, Biometrika, 4, 1906, pp. 332–350.

the regression is linear. In our problem, $N(\eta^2 - r^2) = 104.58$, and the regression is clearly non-linear.

True non-linear relation is often met with in psychophysics, and in experiments dealing with fatigue, practise, forgetting, etc. Most mental and physical tests, however, have been found to exhibit linear relationship, and in consequence r has been employed in psychology and education to a much greater extent than η. If the regression is definitely non-linear, it makes considerable difference whether η or r is taken as the measure of relation. Unless the regression is clearly curvilinear, however, little error is introduced by taking r instead of η; and this is especially true if the correlation is low.

The coefficient of correlation, r, is superior to η in that knowing its value we can easily write the equation from which the value of the dependent variable may be estimated from the independent. This is not possible with the correlation-ratio. In order to estimate one variable from the other in non-linear relation, a curve must be fitted to the means of the arrays of the columns or rows.[1]

X. The Correction of a Coefficient of Correlation for "Attenuation"

The accuracy of any series of test scores or other measures of capacity is always conditioned by the number and size of the chance variations—"errors of observation"—present. The term "errors of observation" may be taken to include slight changes in technique and procedure on the part of the experimenter, as well as variations in the subjects due to fatigue, distraction, shifts in attention or attitude towards the test, and other minor fluctuations of different sorts. If the number of observations is large, errors of observation—since their effect is as liable to be in the negative as the

[1] The subject of curve fitting is fully dealt with in more advanced books on statistics. See, Jones, D. C., *A First Course in Statistics*, 1921, Chaps. XV, XVI, and XVII, for a fairly elementary discussion.

positive direction—will tend in the long run to cancel each other off as far as the average is concerned. Such errors, however, always tend to increase the σ of the distribution, and to decrease or "attenuate" a coefficient of correlation calculated between series in which they are present. For this reason, it is generally advisable to correct raw r's for observational errors, and special formulas have been devised to rule out their effect.[1]

It is first necessary to make at least two independent measures of each capacity, and to find the self-correlation of each test.[2] This done, the r corrected for attenuation may be found from formula (47) given below. The complete procedure is as follows:

Let A and B represent the tests to be correlated.

Let A_1 represent the 1st series of scores obtained in A.

Let A_2 represent the 2nd series of scores obtained in A.

Let B_1 represent the 1st series of scores obtained in B.

Let B_2 represent the 2nd series of scores obtained in B.

Let r_{AB} represent the "true" correlation between tests A and B.

Let $r_{A_1A_2}$ represent the self-correlation of test A.

Let $r_{B_1B_2}$ represent the self-correlation of test B.

Let $r_{A_1B_2}$ represent the obtained correlation between A and B_2.

Let $r_{A_2B_1}$ represent the obtained correlation between A_2 and B_1.

Then [3]

$$r_{AB} = \frac{\sqrt{(r_{A_1B_2})(r_{A_2B_1})}}{\sqrt{(r_{A_1A_2})(r_{B_1B_2})}}, \qquad \ldots\ldots \quad (47)$$

[1] See the two articles by C. Spearman:
(a) *The Proof and Measurement of the Association between Two Things*, American Journal of Psychology, 1904, Vol. XV, p. 72–101.
and (b) *Demonstration of Formulae for True Measure of Correlation*, American Journal of Psychology, 1907, Vol. XVIII, p. 161–169.

[2] See page 288.

[3] See Yule, *An Introduction to the Theory of Statistics*. pp. 213–214 for discussion of this formula.

To illustrate the formula, suppose that A is a Following Directions Test, and B a Mixed Relations Test, and that

$$r_{A_1A_2}=.72 \qquad r_{B_1B_2}=.75$$
$$r_{A_1B_2}=.35 \qquad r_{A_2B_1}=.42$$

Substituting in formula (47) we have

$$r_{AB}=\frac{\sqrt{.35\times .42}}{\sqrt{.72\times .75}}=.52.$$

or correcting for observational errors, we raise the correlation from .35 and .42 (the obtained r's) to .52.

If we have only the one correlation between two given tests A and B, so that formula (47) is inapplicable, it is still possible to obtain an approximate correction for attenuation by dividing the "raw" coefficient by the geometrical mean of the two "reliability coefficients."[1] Formula (47) then becomes

$$r_{AB}=\frac{r_{A_1B_1}}{\sqrt{r_{A_1A_2}r_{B_1B_2}}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (48)$$

Thus if the obtained correlation between tests A and B above had been .50, and the reliability coefficients, as before, .72 and .75, we could correct (approximately) for attenuation as follows:

$$r_{AB}=\frac{.50}{\sqrt{.72\times .75}}=.68.$$

Corrected for attenuation, the obtained coefficient is increased from .50 to .68.

XI. Summary of Formulas Used in This Chapter

1. For Product-Moment r, deviations from GA's

$$r=\frac{\frac{\Sigma xy}{N}-c_xc_y}{\sigma_x\sigma_y}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

[1] See Spearman, C., American Journal of Psychology, 1904, Vol. XV, p. 271.

2. For Product-Moment r, deviations from actual averages

$$r=\frac{\Sigma xy}{N\sigma_x\sigma_y}, \qquad (24)$$

$$r=\frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}, \qquad (25)$$

3. $$PEr=\frac{.6745\times(1-r^2)}{\sqrt{N}}. \qquad (26)$$

4. $$PE_{(\text{diff. } r_1-r_2)}=\sqrt{PE_{r_1}{}^2+PE_{r_2}{}^2}, \qquad (27)$$

5. Regression Equations in Deviation Form

$$y=r\cdot\frac{\sigma_y}{\sigma_x}\cdot x, \qquad (28)$$

$$x=r\cdot\frac{\sigma_x}{\sigma_y}\cdot y, \qquad (29)$$

6. Regression Equations in Score Form

$$Y=r\cdot\frac{\sigma_y}{\sigma_x}(X-X')+Y', \qquad (30)$$

$$X=r\cdot\frac{\sigma_x}{\sigma_y}(Y-Y')+X', \qquad (31)$$

7. Standard Errors of Estimate

$$\sigma_{(\text{est. } Y)}=\sigma_y\sqrt{1-r^2}, \qquad (32)$$

$$\sigma_{(\text{est. } X)}=\sigma_x\sqrt{1-r^2}, \qquad (33)$$

$$PE_{(\text{est. } Y)}=.6745\sigma_y\sqrt{1-r^2}, \qquad (34)$$

$$PE_{(\text{est. } X)}=.6745\sigma_x\sqrt{1-r^2}, \qquad (35)$$

8. Correlation Measured from "Ranks"

$$\rho=1-\frac{6\Sigma D^2}{N(N^2-1)}, \qquad (36)$$

$$PE_r=\frac{.7063(1-r^2)}{\sqrt{N}}, \qquad (37)$$

$$R=1-\frac{6\Sigma G}{(N^2-1)}, \qquad (38)$$

9. Coefficient of Mean Square Contingency, C

$$C=\sqrt{\frac{S-N}{S}}, \quad (39)$$

10. Non-linear Regression

$$\eta_{yx}=\frac{\sigma_{my}}{\sigma_y}, \quad (40)$$

$$PE_\eta=\frac{.6745\times(1-\eta^2)}{\sqrt{N}}, \quad (41)$$

$$\eta_{xy}=\frac{\sigma_{mx}}{\sigma_x}, \quad (42)$$

$$\text{Corrected } \eta^2=\frac{\eta^2-\frac{(\kappa-3)}{N}}{1-\frac{(\kappa-3)}{N}}, \quad (43)$$

$$PE_\zeta= .6745\times 2\sqrt{\frac{\zeta}{N}}\cdot\sqrt{(1-\eta^2)^2-(1-r^2)^2+1}, \quad (44)$$

$$PE_\zeta= .6745\times 2\sqrt{\frac{\zeta}{N}} \text{ (approximately)}, \quad (45)$$

$$N(\eta^2-r^2)<11.37, \quad (46)$$

11. Correction for Attenuation

$$r_{AB}=\frac{\sqrt{(r_{A_1B_2})(r_{A_2B_1})}}{\sqrt{(r_{A_1A_2})(r_{B_1B_2})}}, \quad (47)$$

$$r_{AB}=\frac{r_{A_1B_1}}{\sqrt{(r_{A_1A_2})(r_{B_1B_2})}}, \quad (48)$$

PROBLEMS

1. Find the coefficient of correlation (product-moment) between the following sets of Army Alpha and typewriting scores made by 100 students in a typewriting class. The typewriting scores are

in number of words written per minute (with certain penalties). In tabulating scores, let typing be the Y-variable and Alpha the X-variable. Take the Y-step as 5 and the X-step as 10 units.

Typing (Y)	Alpha (X)	Typing (Y)	Alpha (X)	Typing (Y)	Alpha (X)
46	152	26	164	40	120
31	96	33	127	36	140
46	171	44	144	43	141
40	172	35	160	48	143
42	138	49	106	45	138
41	154	40	95	58	149
39	127	57	146	23	142
46	156	23	175	45	166
34	156	51	126	44	138
48	133	35	120	47	150
48	173	41	154	29	148
38	134	28	146	46	166
26	179	32	154	46	146
37	159	50	159	39	167
34	167	29	175	49	139
51	136	41	164	34	183
47	153	32	111	41	150
39	145	49	164	49	179
32	134	58	119	31	138
37	184	35	160	47	136
26	154	48	149	40	172
40	90	40	149	30	145
53	143	43	143	40	109
46	173	38	159	38	158
39	168	37	157	29	115
52	187	41	153	43	93
47	166	51	149	55	163
31	172	40	163	37	147
33	189	35	175	52	169
22	147	31	133	38	75
46	150	23	178	39	152
44	150	37	168	32	159
37	143	46	156	42	150
31	133				

2. In the Correlation Table [1] given below, find
   (*a*) the coefficient of correlation, and PE_r;
   (*b*) the regression equations in Score Form, and the standard errors of estimate.
   (*c*) What is the most probable height of a boy who weighs 30 pounds? 45 pounds?

[1] See Table XXIV for the C worked out for these data.

Boys: Ages 4.5 to 5.5 Years

Weight in Pounds (X)

Height in Inches (Y)	24–28	29–33	34–38	39–43	44–48	49–53	Totals (Fy)
45–47			1		2		3
42–44			4	35	21	5	65
39–41		5	87	90	7	1	190
36–38	1	18	72	8			99
33–35	5	15	5				25
30–32	2						2
Totals Fx	8	38	169	133	30	6	384

3. In the following correlation table,[1] find
   (*a*) the coefficient of correlation, and the PE_r.
   (*b*) What is the most probable grade of a pupil who makes 120 on Alpha?

Army Alpha *IQ*'s

School Marks	84 and lower	85–89	90–94	95–99	100–104	105–109	110–114	115–119	120–124	125 over	Totals
90 and over				3	3	15	12	9	9	5	56
85–89				8	17	15	24	13	6	6	89
80–84			4	6	22	21	20	10	5	1	89
75–79			7	25	33	23	10	7	4		109
70–74		4	10	18	14	22	12	1	1		82
65–69	1	3	3	12	7	8	8	1			43
60–64			2	5	3	1	1				12
Totals	1	7	26	77	99	105	87	41	25	12	480

[1] From Otis, *Statistical Methods in Educational Measurement*, 1925, p. 315.

4. Find the correlation between the following test scores by
   (*a*) the Rank-Difference Method, and
   (*b*) the Method of Gains.

Individual	Intelligence Score (Alpha)	Cancellation Score (A test + Number Group Checking Test)
Kp	185	110
My	203	98
Le	188	118
Hy	195	104
Sh	176	112
Ld	174	124
Sn	158	119
St	197	95
Wn	176	94
Pe	138	97
Gr	126	110
Bn	160	94
Gm	151	126
Ly	185	120
Ws	185	118

(Note.—Since the Cancellation scores are in seconds, the highest score (94) is numerically the lowest.)

5. Compute the coefficient of contingency, C, for the two tables given below, which show:
   A. The resemblance between brothers in athletic capacity.[1]
   B. The resemblance between fathers and sons in temperament.[2]

A

Athletic Capacity—First Brother

Second Brother	Athletic	Betwixt	Non-athletic	Totals
Athletic	906	20	140	1066
Betwixt	20	76	9	105
Non-athletic	140	9	370	519
Totals	1066	105	519	1690

[1] From Yule, *An Introduction to the Theory of Statistics*, p. 74, after Pearson.
[2] From Brown and Thompson, *Essentials of Mental Measurement*, 1921 p. 125. The coefficient of contingency is not usually calculated for tables having less than a 5×5 fold classification. These tables, however, will illustrate the method in a simple way.

B

FATHERS

Sons	Merry	Melancholy	Alternating	Even	Totals
Merry	122	8	81	67	278
Melancholy	10	2	7	10	29
Alternating	70	9	101	68	248
Even	58	6	66	45	175
Totals	260	25	255	190	730

6. The following correlation table gives the relation between the scores on the Thorndike College Entrance Intelligence Examination and the extra-curricular activities of 102 Columbia College students.[1]

(*a*) Find η_{yx} for this table.

(*b*) Find r, and test the regression of Y on X for linearity.

THORNDIKE SCORES (X)

Extra-curricular Activities (Y)	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90–94	95–99	100–104	F_y
18–20					2	2					4
15–17				2		3	1				6
12–14				4		6	2		2		14
9–11		1	2		4	4	6	7	3		27
6–8	1			6	2	2	6	2	4	1	24
3–5	1		1	3	5	3		5	1	1	20
0–2		1		1			1	1	1	2	7
Totals F_x	2	2	3	16	13	20	16	15	11	4	102

[1] From Sommerville, R. C., *Physical, Motor, and Sensory Traits.* Archives of Psychology, 1924, 75, p. 101.

7. Verify the correlation-ratio η_{xy} of .82 given for Diagram XXVI (see page 209).
   (*a*) Test the regression of X on Y for linearity.
   (*b*) Plot the regression line (or curve) on the diagram.

8. *Ma* is the series of scores from one trial of a memory test.
   Mb is the series of scores from a second trial of the same test.
   Aa is the series of scores from one trial of an association test.
   Ab is the series of scores from a second trial of this test.
   The r's are as follows:

   between M*a* and M*b*, .60.
   between *Mb* and *Aa*, .50.
   between *Ma* and *Ab*, .55.
   between *Aa* and *Ab* .72.

   Find the r between M and A corrected for attenuation.

Answers

1. $r = -.05$; $PEr = .07$.
2. (*a*) $r = .709$; $PEr = .017$.
   (*b*) $Y = .4X + 24.42$; $X = 1.26Y - 11.66$
   $\sigma_{(\text{est. } Y)} = 1.79$; $\sigma_{(\text{est. } X)} = 3.18$.
   (*c*) 36.42 inches; 42.42 inches.
3. (*a*) $r = .455$; $PEr = .024$.
   (*b*) 85.4 with a $PE_{(\text{est. } Y)}$ of 4.75.
4. (*a*) $\rho = .187$; $r = .19$ $PEr = .18$.
   (*b*) $R = .09$; $r = .16$.
5. A. $C = .68$ B. $C = .16$.
6. (*a*) $\eta_{yx} = .43$; η_{yx} (corrected) $= .36$.
   (*b*) $r = -.09$. The regression is almost certainly non-linear.
8. $r = .80$.

CHAPTER V

PARTIAL AND MULTIPLE CORRELATION[1]

I. The Meaning of Partial and Multiple Correlation

The coefficient of correlation between sets of test scores (or other series of measures) often represents not simply the degree of relationship existing between these measures in themselves, but the degree of this relation *plus* the indirect effect of other factors to which they are both related. For this reason in measuring the correlation between two sets of measures, it is necessary that we eliminate or rule out as far as possible those uncontrolled factors which through their common relation to the measures to be correlated tend to raise or lower the "net" correlation. As an illustration of the effect on correlation of uncontrolled factors, suppose that the correlation between intelligence (i) and age (a) in a large group of children whose ages range from 7 to 14 years is r_{ia}; that the correlation between school achievement (s) and age (a) in the same group is r_{sa}; and that the correlation between intelligence (i) and school achievement (s) is r_{is}. Now this last coefficient, r_{is}, is not simply a measure of the influence of intelligence on school achievement, but is a measure of the influence of intelligence, *plus* the *indirect effect of differences in age*, on school achievement. In order to determine the relation between intelligence and school achievement *uninfluenced* by the age factor, it is necessary to rule out the effect of age-differences. This can be accomplished in two ways: (1) by selecting children all of whom are of the same age, or (2) by finding a "partial" coefficient of correlation between intelligence and school

[1] The discussion of partial and multiple correlation given in this chapter follows Yule in method and nomenclature.

standing. Such a partial coefficient is written $r_{is \cdot a}$, and may be thought of as giving the net correlation between intelligence and school achievement for *children of the same age*, or as the net correlation between intelligence and school achievement with *age constant*. In short, a coefficient of partial correlation may be said to represent the net relation between two variables when one or more other variables which might increase or decrease the true correlation have been ruled out or held constant.

In addition to its value as a device whereby we are able to control conditions by ruling out disturbing factors, partial correlation is highly important also in that it enables us to build up regression equations involving three or more variables from which a test score (or other measure) may be predicted when we know the corresponding scores made on the other tests. The value of the regression equation in estimating scores—its accuracy as a predicting instrument—may be determined from the "multiple" coefficient of correlation.[1] This coefficient gives the correlation between the scores actually obtained on a given test, and the scores on the same test predicted by the regression equation from the scores made on two or more *correlated* tests. The multiple coefficient of correlation may be thought of also as giving the correlation between a trait (or traits) as measured by a single test, and the same trait (or traits) as measured by a number of tests taken together. (The multiple coefficient will be best understood by working through an actual problem.)

To summarize briefly, partial and multiple correlation may be considered as representing an important extension of the theory and technique of "simple" or two-variable correlation to include problems which involve three or more variables.

[1] $\sigma_{(est.)}$ also gives the accuracy of the regression equation in predicting single scores. (See page 183.)

II. A Correlation Problem Involving Three Variables

The simplest and most straightforward approach to an understanding of the value of the method of partial and multiple correlation and of the technique involved is by way of an illustration. In the present section, therefore, is shown the application of partial and multiple correlation to a three-variable problem; and following this, the general formulas and some further applications of the method are considered.

The problem selected (Table XXVI) is taken from a study made by Professor Mark May [1] of the factors which influence "academic success." In that part of his study from which our example is taken, May wished to find how accurately he could "predict" the academic success or scholastic achievement of 450 Syracuse freshmen from a knowledge of their general intelligence and study habits. Academic success was defined specifically as the number of "credit" or "honor" points obtained by a student at the end of his first semester in college. The number of honor points secured depends on the number of *A*, *B*, and *C* grades made by the student in his courses. Thus a grade of *A* carries 3 honor points; a grade of *B*, 2 honor points; a grade of *C*, 1 honor point; and a grade of *D*, which is a passing mark, carries no honor point credit. The maximum number of points which a freshman taking the "regular" course can obtain in one semester is 48.

General intelligence was measured by a combination of the Miller Mental Ability Test, and the Dartmouth Completion of Definitions Test. The Miller Test contains 120 items and the Dartmouth Test 40, so that the maximum "raw score" was 160. The scores of the 450 students ranged from 50 to 150, the distribution being fairly normal.

As a measure of industry and application, it was decided to take the number of hours per week spent, on the average, in study. Information in regard to study habits was obtained

[1] May, Mark A., *Predicting Academic Success*, Journal of Educational Psychology, 1923, Vol. XIV, 7, pp. 429–440.

by means of a questionnaire given at the beginning and at the middle of the first semester. Among other items of information asked for in the questionnaire were such things as the number of hours spent per week at meals, in sleeping, etc. In this way an attempt was made to have the student think that he was being checked up on the distribution of his total time, and not on his study habits alone. The self-correlation between the two statements—number of hours spent in study—on the first and second questionnaires was .86, which indicates a very satisfactory degree of reliability.

As previously stated, the main object of this study was to find how accurately the number of honor points which a student receives can be predicted from a knowledge of his study habits and his general intelligence.[1] In solving this problem, however, it is necessary to find the partial coefficient which shows to what extent honor points are related to general intelligence when the variable factor of study-hours per week is held constant; and also the partial coefficient which shows to what extent honor points are related to study-hours when the variable factor of general intelligence is held constant. This information, in itself, will prove to be of considerable interest. The solution of the whole problem is given in the following series of steps—the necessary data and statistics will be found in Table XXVI

Step I. Note that the mean and σ of each series of measures, and the intercorrelations are first calculated. These intercorrelations are the usual product-moment r's, computed as shown in Chapter IV. The r between (1) honor points, and (2) general intelligence, written r_{12} is .60; the r between (1) honor points and (3) number of study hours, written r_{13}, is .32; and the r between (2) general intelligence and (3) number of study hours, i.e., r_{23}, is $-.35$. The low correlation between honor points and study-hours is of considerable interest;

[1] Other factors, of course, such as health, personality, previous preparation, etc., are of considerable importance in determining honor points as May indicates in his article. The two factors selected were chosen simply because they are not only important, but also objective and measurable.

but probably the most interesting r is the $-.35$ between study-hours and general intelligence. Evidently, the brighter the student, the less he studies!

Step II. The next step is to calculate the "net" correlation between (1) honor points and (2) general intelligence with the influence of (3) study-hours "partialed" out or held constant. This net, or partial coefficient of correlation, is written $r_{12.3}$. The formula [1] for $r_{12.3}$ is

$$r_{12.3}=\frac{r_{12}-r_{13}\cdot r_{23}}{\sqrt{1-r^2_{13}}\ \sqrt{1-r^2_{23}}}.\ \text{[Formula (49), page 232].}$$

Substitution of the values of r_{12}, r_{13}, and r_{23} in the formula gives $r_{12.3}$ a value of .802. This means that if all of our 450 students studied exactly the *same number* of *hours per week* (i.e., if the number of study hours were constant), the coefficient of correlation between honor points earned and general intelligence scores would be .802 instead of .60, the obtained coefficient, r_{12}. In other words, if each student spent the same number of hours in study, there would be a much closer correspondence between general intelligence and honor points than there is when the number of study hours varies.

The partial coefficient of correlation between (1) honor points and (3) hours spent in study for (2) general intelligence constant is given by the formula

$$r_{13.2}=\frac{r_{13}-r_{12}r_{23}}{\sqrt{1-r^2_{12}}\sqrt{1-r^2_{23}}}.\qquad \text{[Formula (49)]}$$

Substitution of the values of r_{13}, r_{12} and r_{23} gives a partial coefficient $r_{13.2}=.707$ as against a "raw" coefficient, r_{13}, of .32. It is evident, therefore, that if our group were of *the same degree of general intelligence* [2] there would be a much closer correspond-

[1] The general formulas from which this and other formulas used in this section are derived will be found in Section III following.

[2] By " same degree of general intelligence " is meant the same score on the given general intelligence tests.

ence between the number of honor points received and the number of hours spent in study than there is when the members of the group possess varying degrees of general intelligence—and this is certainly the result to be expected.

The last partial coefficient of correlation $r_{23.1} = -.715$. This coefficient gives the net correlation between (2) general intelligence and (3) study-hours, for (1) honor points held constant, and is found from the formula

$$r_{23.1} = \frac{r_{23} - r_{12} \cdot r_{13}}{\sqrt{1 - r^2_{12}}\sqrt{1 - r^2_{13}}}. \qquad \text{[Formula (49)]}$$

Like the two partial r's above, we may interpret $r_{23.1}$ to mean that the correlation between general intelligence score and hours spent in study in a group in which every student has earned the *same number of honor points* would be much higher—*negatively*—than the raw correlation between these same two factors in a randomly selected group—a group in which the number of honor points received by different students vary. Thus we discover that the brighter students not only study *less* than the average and dull (since $r_{23} = -.35$) but that the *brighter* the student the *less* he *needs* to study in order to reach a given standard of academic success,—to secure a given number of honor points (since $r_{23.1} = -.715$).

Step III. The partial coefficients of correlation calculated, the next step is to write the regression equation from which the most probable number of honor points which a student will receive can be estimated, given his general intelligence score and the number of hours he spends in study per week. The regression equation for three variables is written—in Deviation Form —as follows: [Formula (51)].

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3.$$

In this formula x_1 is the *dependent* variable and stands for honor points; x_2 and x_3 are the independent variables, and

stand for general intelligence and study-hours respectively.[1] In Score Form the equation becomes: [Formula (52)]

$$(X_1 - \text{Av. } X_1) = b_{12.3}(X_2 - \text{Av. } X_2) + b_{13.2}(X_3 - \text{Av. } X_3),$$

or transposing and collecting terms,

$$X_1 = b_{12.3}\, X_2 + b_{13.2}\, X_3 + K \text{ (a constant)}.$$

It is clear that before we can use this equation we must find the values of the *regression coefficients* $b_{12.3}$ and $b_{13.2}$. These are found from the formulas,

$$b_{12.3} = r_{12.3}\frac{\sigma_{1.23}}{\sigma_{2.13}}; \quad \text{and} \quad b_{13.2} = r_{13.2}\frac{\sigma_{1.23}}{\sigma_{3.12}}, \quad \text{[Formula (53)]}$$

and as we already have the value of $r_{12.3}$ and $r_{13.2}$ it is only necessary to find $\sigma_{1.23}$, $\sigma_{2.13}$, and $\sigma_{3.12}$ (the "partial" σ's) in order to replace the regression coefficients in the equation by numerical values.

Step IV. The values of the "partial" σ's are found from the formulas,

1. $\sigma_{1.23} = \sigma_1\sqrt{1 - r^2_{12}}\sqrt{1 - r^2_{13.2}}$.
2. $\sigma_{2.13} = \sigma_2\sqrt{1 - r^2_{23}}\sqrt{1 - r^2_{12.3}}$. [Formula (50)]
3. $\sigma_{3.12} = \sigma_3\sqrt{1 - r^2_{23}}\sqrt{1 - r^2_{13.2}}$.

Substituting the known values of the raw and partial r's in these formulas we get $\sigma_{1.23} = 6.34$; $\sigma_{2.13} = 8.84$; $\sigma_{3.12} = 3.97$. (For calculations, see Table XXVI.)

Step V. From the partial σ's and the partial r's, the numerical values of the regression coefficients $b_{12.3}$ and $b_{13.2}$ are found to be .57 and 1.13, respectively. Hence we may now write the regression equation as

$$x_1 = .57x_2 + 1.13x_3;$$

or multiplying by a convenient constant (e.g., by 1.75), (the number of honor points) = 1 (score on the intelligence tests) + 2 (number of hours spent in study per week). It is evident from this equation that in so far as the general intelligence score and

[1] Note the resemblance of this equation to the simple regression equation for two variables $y = b_{12} \cdot x$ (page 174). If x_1 is put for y and x_2 for x in this equation, we have, $x_1 = b_{12} \cdot x_2$.

number of study hours per week determine the number of honor points received, their relative weight is as 1 : 2.

TABLE XXVI

A Correlation Problem Involving Three Variables

Step I

(1) Honor Points	(2) General Intelligence	(3) Hours of Study per Week
$M_1 = 18.5$	$M_2 = 100.6$	$M_3 = 24$
$\sigma_1 = 11.2$	$\sigma_2 = 15.8$	$\sigma_3 = 6$
$r_{12} = .60$	$r_{13} = .32$	$r_{23} = -.35$

Step II. Calculation of Partial Coefficients of Correlation. (see Note)

$$r_{12.3}{}^* = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r^2_{13}}\sqrt{1-r^2_{23}}} = \frac{.60 - .32(-.35)}{.9474 \times .9367} = .802. \quad . \quad . \quad (49)$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1-r^2_{12}}\sqrt{1-r^2_{23}}} = \frac{.32 - .60(-.35)}{.8 \times .9367} = .707.$$

$$r_{23.1} = \frac{r_{23} - r_{12}r_{.3}}{\sqrt{1-r^2_{12}}\sqrt{1-r^2_{13}}} = \frac{-.35 - .32 \times .60}{.8 \times .9474} = -.715.$$

* For $\sqrt{1-r^2}$ values, use Table XXVII.

Step III. The Regression Equations

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3 \quad \text{(Deviation Form)}, \quad . \quad . \quad . \quad . \quad (51)$$

or

$$X_1 = b_{12.3}X_2 + b_{13.2}X_3 + K. \quad \text{(Score Form)}, \quad . \quad . \quad . \quad . \quad (52)$$

in which

$$b_{12.3} = r_{12.3}\frac{\sigma_{1.23}}{\sigma_{2.13}} \quad \text{and} \quad b_{13.2} = r_{13.2}\frac{\sigma_{1.23}}{\sigma_{3.12}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

Step IV. Calculation of σ's

(1) $\sigma_{1\,23} = \sigma_1\sqrt{1-r^2_{12}}\sqrt{1-r^2_{13.2}} = 11.2 \times .8 \times .7072 = 6.34.$. (50)

(2) $\sigma_{2.13} = \sigma_2\sqrt{1-r^2_{23}}\sqrt{1-r^2_{12.3}} = 15.8 \times .9367 \times .5973 = 8.84$

(3) $\sigma_{3.12} = \sigma_3\sqrt{1-r^2_{23}}\sqrt{1-r^2_{13.2}} = 6 \times .9367 \times .7072 = 3.97$

Step V. The Regression Coefficients and Regression Equation

Substituting for $r_{12.3}$, $r_{13.2}$, $\sigma_{1.23}$, $\sigma_{2.13}$, $\sigma_{3.12}$, we have

$$b_{12.3} = .802 \times \frac{6.34}{8.84} = .57; \quad b_{13.2} = .707 \times \frac{6.34}{3.97} = 1.13.$$

Hence the regression equation becomes:

$$x_1 = .57x_2 + 1.13x_3 \quad \text{(Deviation Form)},$$

or

$$X_1 = .57X_2 + 1.13X_3 - 66 \quad \text{(Score Form)}.$$

Step VI. Calculation of the Standard Error of Estimate

$$\sigma_{(\text{est. } X_1)} = \sigma_{1.23} = 6.34. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

$$PE_{(\text{est. } X_1)} = .6745 \times 6.34 = 4.28. \quad . \quad . \quad . \quad . \quad . \quad . \quad (55)$$

Step VII. The Coefficient of Multiple Correlation

$$R_1(23) = \sqrt{1 - \frac{\sigma^2_{1.23}}{\sigma^2_1}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (56)$$

$$= .824$$

Note.—It should be noted that while the partial coefficient of correlation $r_{23.1}$ is of interest as giving us the relation between general intelligence and hours

spent in study for a constant number of honor points, it is unnecessary in the regression equation, $x_1 = b_{12.3}x_2 + b_{13.2}x_3$. In order to evaluate the constants $b_{12.3}$ and $b_{13.2}$ in this regression equation, we need only $r_{12.3}$ and $r_{13.2}$. In any problem involving three variables, only two partial coefficients of correlation need be computed, if we are interested only in the prediction of X_1 values from known values of X_2 and X_3.

TABLE XXVII

A TABLE TO INFER THE VALUE OF $\sqrt{1-r^2}$ FROM A GIVEN VALUE OF r

r	$\sqrt{1-r^2}$	r	$\sqrt{1-r^2}$	r	$\sqrt{1-r^2}$
.00	1.0000	.34	.9404	.68	.7332
.01	.9999	.35	.9367	.69	.7238
.02	.9998	.36	.9330	.70	.7141
.03	.9995	.37	.9290	.71	.7042
.04	.9992	.38	.9250	.72	.6940
.05	.9987	.39	.9208	.73	.6834
.06	.9982	.40	.9165	.74	.6726
.07	.9975	.41	.9121	.75	.6614
.08	.9968	.42	.9075	.76	.6499
.09	.9959	.43	.9028	.77	.6380
.10	.9950	.44	.8980	.78	.6258
.11	.9939	.45	.8930	.79	.6131
.12	.9928	.46	.8879	.80	.6000
.13	.9915	.47	.8827	.81	.5864
.14	.9902	.48	.8773	.82	.5724
.15	.9887	.49	.8717	.83	.5578
.16	.9871	.50	.8660	.84	.5426
.17	.9854	.51	.8617	.85	.5268
.18	.9837	.52	.8542	.86	.5103
.19	.9818	.53	.8480	.87	.4931
.20	.9798	.54	.8417	.88	.4750
.21	.9777	.55	.8352	.89	.4560
.22	.9755	.56	.8285	.90	.4359
.23	.9732	.57	.8216	.91	.4146
.24	.9708	.58	.8146	.92	.3919
.25	.9682	.59	.8074	.93	.3676
.26	.9656	.60	.8000	.94	.3412
.27	.9629	.61	.7924	.95	.3122
.28	.9600	.62	.7846	.96	.2800
.29	.9570	.63	.7766	.97	.2431
.30	.9539	.64	.7684	.98	.1990
.31	.9507	.65	.7599	.99	.1411
.32	.9474	.66	.7513	1.00	.0000
.33	.9440	.67	.7424		

To write the regression in Score Form, we simply replace x_1 by $(X_1 - 18.5)$; x_2 by $(X_2 - 100.6)$; and x_3 by $(X_3 - 24)$. The equation then becomes

$$X_1 = .57X_2 + 1.13X_3 - 66.$$

Given a student's general intelligence score (X_2) and the number of hours he spends in study per week (X_3) we can, from this equation, estimate the most probable number of honor points which he will receive in the first semester. By way of illustration, suppose that a student has a general intelligence score of 120 points and that he studies on the average 20 hours per week: how many honor point will he most probably receive during the first semester? Substituting $X_2 = 120$ and $X_3 = 20$ in the regression equation, we have that

$$X_1 = .57 \times 120 + 1.13 \times 20 - 66, \text{ or } X_1 = 25.$$

The most probable number of honor points which this student will receive, therefore, using the given criteria as the basis of our estimate, is 25.

Step VI. This estimate—like every other "most probable" number of honor points predicted from the regression equation —has a certain "error of estimate." The standard error of estimate of all honor points, i.e., X_1's, predicted from the regression equation $X_1 = b_{12.3}X_2 + b_{13.2}X_3 + K$ is designated $\sigma_{(\text{est. } X_1)}$ and equals $\sigma_{1.23}$ [see Formula (50)] directly. The $PE_{(\text{est. } X_1)}$ is $.6745 \times \sigma_{(\text{est. } X_1)}$.

The standard error of estimate in the present problem is 6.34 points, and the $PE_{(\text{est. } X_1)}$ is 4.28 points. In the illustration above, therefore, the 25 estimated honor points have a $PE_{(\text{est. } X_1)}$ of 4.28 points, which means that the chances are even—50 in 100—that this student will receive (roughly) not less than 21 nor more than 29 honor points. The reliability of any other honor points estimate made from the regression equation may be found in exactly the same way.

Step VII. The final step in the solution of our problem is to compute the coefficient of multiple correlation. This "multiple r," which is generally written R[1], has been defined (see page 222) as the coefficient of correlation between the scores

[1] Multiple R must not be confused with the R of the Spearman Footrule formula, page 194.

actually made on a given test and the scores on the same test predicted from the regression equation. Expressed more mathematically, R gives the correlation between the *dependent* variable X_1, and the *independent* variables, X_2, X_3, etc., taken together as a team. The formula for R when there are two independent variables is

$$R_{1(23)} = \sqrt{1 - \frac{\sigma^2_{1.23}}{\sigma^2_1}}. \qquad \text{[Formula (56)]}$$

In the present problem, $R_{1(23)} = .824$. This means that if the most probable number of honor points which each student in our group of 450 will receive is predicted from the regression equation, the correlation between these 450 *predicted* scores and the 450 scores *actually received* will be .824. Multiple R, therefore, tells us how closely X_1 is related to the *combined action* of X_2 and X_3, or—in the present instance—how closely honor points are related to general intelligence and number of hours spent in study per week, taken together.

III. General Formulas for Use in Partial and Multiple Correlation

I. General Formulas for Partial *r*'s

We have found (Table XXVI) that in a correlation problem involving three variables, we are enabled by the method of partial correlation to find the net relation between *two* variables when a *third* is ruled out or held constant. In like manner, by an extension of the method of partial correlation, we can secure the net correlation between X_1 and X_2 when two or more variables have been ruled out or held constant. Thus the partial coefficient of correlation $r_{12.34}$ means by analogy to $r_{12.3}$ that the correlation between X_1 and X_2 has been freed of the influence of both X_3 and X_4; and the partial coefficient of correlation $r_{12.34 \ldots n}$ means that the correlation between X_1 and X_2 has been freed (theoretically) of the influence of all disturbing factors.

In every partial coefficient of correlation the subscripts to the *left* of the point are called *primary* subscripts and denote the two variables whose correlation we are seeking. The subscripts to the *right* of the point are called *secondary* subscripts, and denote those variables which are to be ruled out or held constant.[1] The order of a partial r is determined by the number of its secondary subscripts: $r_{12.3}$ or $r_{13.2}$ or $r_{23.1}$, for example, is a partial r of the first order, while "entire" or "total" r's, such as r_{12} or r_{13} or r_{23} are coefficients of zero order.

The general formula for partial r's of the nth order is written

$$r_{12.34\ldots n}=\frac{r_{12.34\ldots(n-1)}-r_{1n.34\ldots(n-1)}r_{2n.34\ldots(n-1)}}{\sqrt{1-r^2_{1n.34\ldots(n-1)}}\sqrt{1-r^2_{2n.34\ldots(n-1)}}}. \qquad (49)$$

From formula (49) partial r's of any given order can be found. In a four-variable problem, for example, $r_{12.34}$ may be written by reference to the formula as

$$r_{12.34}=\frac{r_{12.3}-r_{14.3}r_{24.3}}{\sqrt{1-r^2_{14.3}}\sqrt{1-r^2_{24.3}}},$$

that is to say, in terms of the partial r's of the *first* order. These first order partial r's must then be computed by (49) from r's of zero order before the second order r's can be evaluated. To find partial r's of a higher order, we must first express them in terms of the partial r's of the next lower order; and these r's, in turn, in terms of r's of the next lower order, and so on until r's of zero order have been reached.[2] In other words, it is necessary to "work up" from zero order r's, whenever r's of any higher order are to be computed. Hence it is apparent that with each additional variable the arithmetic of calculation

[1] The order in which the secondary subscripts are written is entirely immaterial, e.g., $r_{12.34}=r_{12.43}$. The order of the primary subscripts is of importance, however, in telling us which variable is "dependent" and which "independent." Thus r_{12} means that X_1 is dependent—is to be predicted from X_2; while r_{21} means that X_2 is dependent—is to be predicted from X_1. The numerical value of r_{12} and r_{21} is, of course, the same.

[2] In calculating partial r's, use Table XXVII to get $\sqrt{1-r^2}$ values.

is greatly increased. As a result, unless the work is carefully planned, the calculations soon become extremely laborious.

The *PE* of a partial r of any order may be found, like the *PE* of an "entire" r, by substituting in formula (26).

2. General Formulas for Partial σ's of Any Order

Just as the correlation between two sets of scores or other measures can be determined when the influence of 1, 2, 3, . . . n other factors is held constant, so the variability (the σ) of any set of scores can be found when the influence of 1, 2, 3, . . . n factors is held constant. As an illustration of this, take $\sigma_{1.23}$ of Table XXVI. This "partial σ" gives the variability of X_1 (honor points) freed of the influence exerted by the two factors X_2 (general intelligence) and X_3 (average study-hours per week). The general formula for σ's of any order is

$$\sigma_{1.234\ldots n}=\sigma_1\sqrt{1-r^2_{12}}\sqrt{1-r^2_{13.2}}\sqrt{1-r^2_{14.23}}\ldots\sqrt{1-r^2_{1n.23\ldots(n-1)}} \qquad (50)$$

This formula may be used to compute the net σ's in correlation problems which involve any number of variables. In a five-variable problem, for example, $\sigma_{1.2345}$ is written

(1) $\sigma_{1.2345}=\sigma_1\sqrt{1-r^2_{12}}\sqrt{1-r^2_{13.2}}\sqrt{1-r^2_{14.23}}\sqrt{1-r^2_{15.234}}$

and by analogy to (1) or by reference to (50) the other σ's may be written:

(2) $\sigma_{2.1345}=\sigma_2\sqrt{1-r^2_{12}}\sqrt{1-r^2_{23.1}}\sqrt{1-r^2_{24.13}}\sqrt{1-r^2_{25.134}}$

(3) $\sigma_{3.1245}=\sigma_3\sqrt{1-r^2_{13}}\sqrt{1-r^2_{23.1}}\sqrt{1-r^2_{34.12}}\sqrt{1-r^2_{35.124}}$

(4) $\sigma_{4.1235}=\sigma_4\sqrt{1-r^2_{14}}\sqrt{1-r^2_{24.1}}\sqrt{1-r^2_{34.12}}\sqrt{1-r^2_{45.123}}$

(5) $\sigma_{5.1234}=\sigma_5\sqrt{1-r^2_{15}}\sqrt{1-r^2_{25.1}}\sqrt{1-r^2_{35.12}}\sqrt{1-r^2_{45.123}}$

Each of these σ's measures the variability of a *single* factor when the effects of the other *four* are ruled out or held constant. All of them are σ's of the fourth order, since there are 4 secondary subscripts, and the order of a partial σ, like the order

of a partial r, is determined by the number of its secondary subscripts.

By a simple rearrangement of the secondary subscripts any higher order σ may be written in more than one way. A σ of the second order may be written in two ways: e.g., $\sigma_{1.23}$ which is given on page 227 as $\sigma_{1.23}=\sigma_1\sqrt{1-r^2_{12}}\sqrt{1-r^2_{13.2}}$ may also be written $\sigma_{1.32}=\sigma_1\sqrt{1-r^2_{13}}\sqrt{1-r^2_{12.3}}$.
In like manner, $\sigma_{2.13}$ may be written

$$(1)\quad \sigma_{2.13}=\sigma_2\sqrt{1-r^2_{12}}\sqrt{1-r^2_{23.1}},$$

or

$$(2)\quad \sigma_{2.31}=\sigma_2\sqrt{1-r^2_{23}}\sqrt{1-r^2_{12.3}};$$

and $\sigma_{3.12}$ may be written

$$(1)\quad \sigma_{3.12}=\sigma_3\sqrt{1-r^2_{13}}\sqrt{1-r^2_{23.1}}$$

or

$$(2)\quad \sigma_{3.21}=\sigma_3\sqrt{1-r^2_{23}}\sqrt{1-r^2_{13.2}}.$$

The alternate forms of a partial σ are useful as a check on the arithmetic calculations, and too because they make unnecessary the calculation of otherwise unused and hence superfluous partial r's. Thus by using the *second forms* of $\sigma_{2.13}$ and $\sigma_{3.12}$ instead of the *first* (see Table XXVI) we make unnecessary the calculation of $r_{23.1}$ so far as the computation of the σ's is concerned. Furthermore, if $r_{23.1}$ is not used elsewhere in the problem, it need not be calculated at all (see page 228). Two partial r's, are all that we need in order to write the regression equation in a three-variable problem.

The number of alternate forms in which any higher order σ may be written depends on the number of permutations which its secondary subscripts can take. We have seen that a second order σ may be written in two ways: $\sigma_{1.23}$ and $\sigma_{1.32}$. In the same way, any σ of the third order, e.g., $\sigma_{1.234}$ may be written in 6 ways: $\sigma_{1.234}$, $\sigma_{1.243}$, $\sigma_{1.324}$, $\sigma_{1.342}$, $\sigma_{1.423}$, $\sigma_{1.432}$. Any σ of the fourth order, e.g., $\sigma_{1.2345}$ may be written in 24 ways, and any σ of the fifth order, e.g., $\sigma_{1.23456}$, in 120 ways.[1]

[1] This follows from the law of permutations. The permutations of 4 things taken 4 at a time are ${}_4P_4=4\times3\times2\times1=24$; and the permutations of 5 things

Fortunately we need only a very few of all of these possible arrangements. Care, nevertheless, must be taken that the correct forms are chosen, for just as the number of partial r's which must be computed in a 3-variable problem can be reduced by a judicious choice of σ formulas, so also in problems which contain more than 3 variables the number of partial r's may be considerably reduced by proper selection. And it is in the longer problems that a reduction of the number of partial r's to be computed counts most, since it is here that the calculations become laborious. The partial σ's which require the calculation of the minimum number of partial r's are given—for 4- and 5-variable problems—in the outline solutions on pages 240–244. These will be found useful for quick reference. By analogy to these, the selection of the σ formulas in problems which involve more than five variables can be easily made.

3. General Formulas for the Regression Equation, and Coefficients of Regression

The general regression equation, which expresses the relation between a single dependent variable, X_1, and a number of independent variables, $X_2, X_3, X_4 \ldots X_n$, may be written in Deviation Form as follows:

$$x_1 = b_{12.34\ldots n}\, x_2 + b_{13.24\ldots n}\, x_3 + \ldots b_{1n.23\ldots(n-1)}\, x_n. \quad (51)$$

and in Score Form as

$$X_1 = b_{12.34\ldots n} X_2 + b_{13.24\ldots n} X_3 + \ldots b_{1n.23\ldots(n-1)} X_n + K. \quad (52)$$

The regression coefficients $b_{12.34\ldots n}$, $b_{13.24\ldots n}$, etc., give the *weight* or value to be attached to each independent variable when X_1 is to be estimated from all of these in combination. Moreover, the regression coefficients indicate the weight which *each independent variable* has in determining X_1 *exclusive* of the influence of the other variables, and hence we can tell from the regression equation just what part the score on each of several

taken 5 at a time are ${}_5P_5 = 5 \times 4 \times 3 \times 2 \times 1 = 120$. In general, the permutations of n things taken n at a time are ${}_nP_n = n(n-1)(n-2)\ldots$ to n factors. See the Chapter on Permutations and Combinations in any Algebra.

tests plays in determining the score on the test taken as the dependent variable.

The regression coefficients in a regression equation may be computed from the formula

$$b_{12.34\ldots n} = r_{12.34\ldots n}\frac{\sigma_{1.234\ldots n}}{\sigma_{2.134\ldots n}}. \quad . \quad . \quad . \quad (53)$$

If the problem involves only three variables, the regression equation becomes $X_1 = b_{12.3}X_2 + b_{13.2}X_3 + K$. In this equation, the regression coefficients $b_{12.3}$ and $b_{13.2}$ are—like the partial r's, $r_{12.3}$, and $r_{13.2}$—of the first order. The first, $b_{12.3}$, equals $r_{12.3}\frac{\sigma_{1.23}}{\sigma_{2.13}}$; and the second, $b_{13.2}$, equals $r_{13.2}\frac{\sigma_{1.23}}{\sigma_{3.12}}$ (see page 227 and Table XXVI). Regression equations which involve more than three variables are easily written by reference to formula (52) and their regression coefficients may be found from formula (53). In a five-variable problem, for example, the regression equation becomes

$$X_1 = b_{12.345}X_2 + b_{13.245}X_3 + b_{14.235}X_4 + b_{15.234}X_5 + K,$$

and the regression coefficients (b's of the third order) are

$$b_{12.345} = r_{12.345}\frac{\sigma_{1.2345}}{\sigma_{2.1345}}$$

$$b_{13.245} = r_{13.245}\frac{\sigma_{1.2345}}{\sigma_{3.1245}}$$

$$b_{14.235} = r_{14.235}\frac{\sigma_{1.2345}}{\sigma_{4.1235}}$$

$$b_{15.234} = r_{15.234}\frac{\sigma_{1.2345}}{\sigma_{5.1234}}$$

Obviously, to compute these regression coefficients we must first compute the third order partial r's, and the necessary partial σ's. The calculation of the b's is then a matter of substitution.

4. General Formulas for Standard and Probable Errors of Estimate

All X_1 scores estimated from a regression equation have a standard error of estimate, $\sigma_{(\text{est. } X_1)}$, which measures the error made in taking estimated instead of actual scores (see page 230). $\sigma_{(\text{est. } X_1)}$ is found from the formula for $\sigma_{1.234 \ldots n}$, as follows:

$$\sigma_{(\text{est. } X_1)} = \sigma_{1.234 \ldots n}, \qquad (54)$$

and

$$PE_{(\text{est. } X_1)} = .6745 \times \sigma_{(\text{est. } X_1)}. \qquad (55)$$

As $\sigma_{1.234 \ldots n}$ must always be computed in order to find the regression coefficients (see examples above), $\sigma_{(\text{est. } X_1)}$ is known at once without further calculation. The value of a standard error of estimate has already been illustrated on page 230 from the data of Table XXVI. To repeat, we find in Table XXVI, that the $\sigma_{(\text{est. } X_1)}$ of any estimated number of honor points is 6.34, and that the $PE_{(\text{est. } X_1)}$ is 4.28 points. Hence, the chances are even that the "most probable," i.e., estimated, number of honor points received by any student—as found from the regression equation—will be in error by 4 points or less (roughly). We may be practically certain that any estimated number of honor points is not in error by *more* than 4×4 or 16 honor points.

It may be shown by the method of least squares [1] that the standard error (or PE) of estimate is a *minimum* when the regression equation is used to estimate the X_1 scores. For this reason, values of X_1 predicted from the regression equation are said to be the "best" estimates of the actual X_1 values which can be made from a linear equation which contains the given variables. The regression equation $X_1 = .57X_2 + 1.13X_3 - 66$ (see page 230) will serve as an illustration of what is meant. Assuming that the relation between X_1 and X_2, X_1 and X_3, and X_2 and X_3 is *linear* in every case, X_1 (honor points) can be estimated from this equation with a *smaller* error of estimate than from any other equation.

[1] See Yule, *An Introduction to the Theory of Statistics*, p. 231.

5. General Formula for R, the Coefficient of Multiple Correlation

The correlation between a single dependent variable X_1 and $(n-1)$ independent variables,—e.g., X_2, X_3, $X_4 \ldots X_n$—in combination is given by the formula

$$R_{1(23 \ldots n)} = \sqrt{1 - \frac{\sigma^2_{1.23 \ldots n}}{\sigma^2_1}}, \quad . \quad . \quad . \quad . \quad (56)$$

in which $R_{1(23 \ldots n)}$ is the coefficient of multiple correlation, σ_1 is the σ of the dependent series of X_1 scores, and $\sigma_{1.23 \ldots n}$ equals the standard error of estimate (see formula 54). When there are only three variables, the multiple coefficient of correlation becomes $R_{1(23)} = \sqrt{1 - \frac{\sigma^2_{1.23}}{\sigma^2_1}}$; when there are five variables $R_{1(2345)} = \sqrt{1 - \frac{\sigma^2_{1.2345}}{\sigma^2_1}}$; and in like manner the R for six, seven, or any number of variables may be written by reference to (56).

Since the error of estimate is a *minimum* when the regression equation is used for estimating X_1 scores, it follows that the multiple coefficient of correlation R gives the maximum correlation obtainable between the *actual* X_1 scores and X_1 scores *estimated* from a knowledge of the independent variables X_2, $X_3 \ldots X_n$, in the regression equation. R is valuable, therefore, as indicating how effectively a given combination of measures (or "team of tests") represents the actual values of X_1 when these measures are combined in the best possible way. R is always positive no matter what the signs in the regression equation may be. Errors of sampling, therefore, do not neutralize each other but tend to become *cumulative*. As a result, the PE of R—which is found from the same formula as the PE of any product-moment r—is not a fair measure of the coefficient's validity. To test the validity of an obtained R, we must compare it with the value of that R which we should get from the same number of cases and the same number of variables, when the variables are uncorrelated,

i.e., with the R which would arise from fluctuations of sampling alone. The formula for this R is

$$R=\frac{\sqrt{n-1}}{\sqrt{N}}, \quad \ldots \ldots \ldots \quad (57)$$

in which n is the number of variables, and N is the number of cases.[1] To illustrate this formula, let us apply it to the three-variable problem in Table XXIV, in which $n=3$, and $N=450$. Substituting for N and n in the formula, we get an R equal to .07, which indicates a highly satisfactory degree of validity for the obtained R of .824.

If we replace $\sigma_{1.23\ldots\ldots n}$ in formula (56) by its value in terms of the entire and partial r's [see formula 50] we may write the general formula for $R_{1(234\ldots n)}$, as follows:

$$R_{1(234\ldots n)}=\sqrt{1-[(1-r^2_{12})(1-r^2_{13.2})\ldots(1-r^2_{1n.23\ldots(n-1)})]}. \quad . \quad (58)$$

Moreover, since a higher order σ may be written in a variety of ways, the number depending upon its order (see page 234), we have in the alternate forms for R a valuable means of checking the accuracy of our arithmetical calculations. In a three-variable problem, for example, $R_{1(23)}$ may be written as

$$R_{1(23)}=\sqrt{1-[(1-r^2_{12})(1-r^2_{13.2})]},$$

or

$$R_{1(32)}=\sqrt{1-[(1-r^2_{13})(1-r^2_{12.3})]}.$$

In like manner, in a 4-variable problem $R_{1(234)}$ may be found from

$$R_{1(234)}=\sqrt{1-[(1-r^2_{12})(1-r^2_{13.2})(1-r^2_{14.23})]},$$

and checked by

$$R_{1(342)}=\sqrt{1-[(1-r^2_{13})(1-r^2_{14.3})(1-r^2_{12.34})]}.$$

[1] Rosenow, Curt, *The Analysis of Mental Functions*, Psychological Monographs, 1917, Vol. XXIV, 5, p. 20.

6. Outline of the Formulas Needed in Correlation Problems Which Involve (a) Four Variables and (b) Five Variables

In multiple correlation problems, generally the main task is to find—with a minimum of time and calculation—the regression equation which expresses the relation of the dependent variable to the independent variables. For this purpose, when working with more than three variables, the simplest plan is to write down the formula for the regression equation required *first* and then proceed *deductively* to find those partial r's and higher order σ's which are necessary for computing the regression coefficients. The formulas for getting the regression equation with a minimum amount of calculation are given—for four and five variables—in the following outlines. It is necessary, of course, that all zero order r's be first computed before the partial correlation technique can be applied.

(a) Formulas for Four-Variable Problems

(1) Regression Equation. The regression equation for four variables is written by reference to formula (52) as follows:

$$X_1 = b_{12.34}X_2 + b_{13.24}X_3 + b_{14.23}X_4 + K.$$

(2) Regression Coefficients. The three regression coefficients needed in (1) are found from formula (53),—

$$b_{12.34} = r_{12.34}\frac{\sigma_{1.234}}{\sigma_{2.134}}$$

$$b_{13.24} = r_{13.24}\frac{\sigma_{1.234}}{\sigma_{3.124}}$$

$$b_{14.23} = r_{14.23}\frac{\sigma_{1.234}}{\sigma_{4.123}}$$

These regression coefficients evidently require the computation of 3 second order partial r's, and 4 third order σ's.

(3) Partial r's.

To find:

(a)

$$r_{12.34}=\frac{r_{12.3}-r_{14.3}\,r_{24.3}}{\sqrt{1-r^2_{14.3}}\sqrt{1-r^2_{24.3}}}$$

We must find 3 first order partial r's as follows:

$$r_{12.3}=\frac{r_{12}-r_{13}\,r_{23}}{\sqrt{1-r^2_{13}}\sqrt{1-r^2_{23}}}$$

$$r_{14.3}=\frac{r_{14}-r_{13}\,r_{34}}{\sqrt{1-r^2_{13}}\sqrt{1-r^2_{34}}}$$

$$r_{24.3}=\frac{r_{24}-r_{23}\,r_{34}}{\sqrt{1-r^2_{23}}\sqrt{1-r^2_{34}}}$$

To find:

(b)

$$r_{13.24}=\frac{r_{13.2}-r_{14.2}\,r_{34.2}}{\sqrt{1-r^2_{14.2}}\sqrt{1-r^2_{34.2}}}$$

We must find 3 first order partial r's as follows:

$$r_{13.2}=\frac{r_{13}-r_{12}\,r_{23}}{\sqrt{1-r^2_{12}}\sqrt{1-r^2_{23}}}$$

$$r_{14.2}=\frac{r_{14}-r_{12}\,r_{24}}{\sqrt{1-r^2_{12}}\sqrt{1-r^2_{24}}}$$

$$r_{34.2}=\frac{r_{34}-r_{23}\,r_{24}}{\sqrt{1-r^2_{23}}\sqrt{1-r^2_{24}}}$$

To find:

(c)

$$r_{14.23}=\frac{r_{14.2}-r_{13.2}\,r_{34.2}}{\sqrt{1-r^2_{13.2}}\sqrt{1-r^2_{34.2}}}$$

No partials of first order are needed other than those already found.

[Note that a minimum of 9 partial r's must be computed, 3 of the second order and 6 of the first order. The 9 first and second order r's together with the 6 zero order r's make 15 coefficients of correlation required in all.]

(4) Standard Deviations. The four third order σ's required may be found from the following formulas which make use of no partial r's other than those already computed in **(3)** above. From formula (50):

$$\sigma_{1.234}=\sigma_1\sqrt{1-r^2_{12}}\sqrt{1-r^2_{13.2}}\sqrt{1-r^2_{14.23}}$$

$$\sigma_{2.134}\ (\text{i.e., }\sigma_{2.341})=\sigma_2\sqrt{1-r^2_{23}}\sqrt{1-r^2_{24.3}}\sqrt{1-r^2_{12.34}}$$

$$\sigma_{3.124}\ (\text{i.e., }\sigma_{3.241})=\sigma_3\sqrt{1-r^2_{23}}\sqrt{1-r^2_{34.2}}\sqrt{1-r^2_{13.24}}$$

$$\sigma_{4.123}\ (\text{i.e., }\sigma_{4.321})=\sigma_4\sqrt{1-r^2_{34}}\sqrt{1-r^2_{24.3}}\sqrt{1-r^2_{14.23}}$$

The numerical values of the regression coefficients may now be computed and substituted in the regression equation.

(5) The Standard Error of Estimate, $\sigma_{(\text{est. }X_1)}$. From formulas (54) and (55) we find:

$$\sigma_{(\text{est. }X_1)}=\sigma_{1.234}\qquad[\text{for value }\sigma_{1.234}\text{ see }(\mathbf{4})\text{ above}]$$

$$PE_{(\text{est. }X_1)}=.6745\,\sigma_{(\text{est. }X_1)}$$

(6) **Coefficient of Multiple Correlation, R.** In a four-variable problem the multiple coefficient, R, is written $R_{1(234)}$ and may be found from formula (56):

$$R_{1(234)} = \sqrt{1 - \frac{\sigma^2_{1.234}}{\sigma_1^2}}$$

This formula may also be written as:

$$R_{1(234)} = \sqrt{1 - [(1 - r^2_{12})(1 - r^2_{13.2})(1 - r^2_{14.23})}$$

or as

$$R_{1(234)} = \sqrt{1 - [(1 - r^2_{13})(1 - r^2_{14.3})(1 - r^2_{12.34})}$$

(*b*) Formulas for Five-Variable Problems

(1) Regression Equation:

$$X_1 = b_{12.345}X_2 + b_{13.245}X_3 + b_{14.235}X_4 + b_{15.234}X_5 + K. \quad . \quad . \quad (52)$$

(2) Regression Coefficients:

$$b_{12.345} = r_{12.345}\frac{\sigma_{1.2345}}{\sigma_{2.1345}}, \qquad b_{14.235} = r_{14.235}\frac{\sigma_{1.2345}}{\sigma_{4.1235}}, \quad . \quad . \quad (53)$$

$$b_{13.245} = r_{13.245}\frac{\sigma_{1.2345}}{\sigma_{3.1245}}, \qquad b_{15.234} = r_{15.234}\frac{\sigma_{1.2345}}{\sigma_{5.1234}}.$$

(3) Partial r's. We compute 22 partial r's as follows (formula 49):

(*a*)

To find: $r_{12.345}$ write as $r_{12.453}$. Then—

$$r_{12.453} = \frac{r_{12.45} - r_{13.45}\,r_{23.45}}{\sqrt{1 - r^2_{13.45}}\sqrt{1 - r^2_{23.45}}}.$$

To compute this r we need 3 partial r's of the second order, viz.,—

$$r_{12.45} = \frac{r_{12.4} - r_{15.4}\,r_{25.4}}{\sqrt{1 - r^2_{15.4}}\sqrt{1 - r^2_{25.4}}},$$

$$r_{13.45} = \frac{r_{13.4} - r_{15.4}\,r_{35.4}}{\sqrt{1 - r^2_{15.4}}\sqrt{1 - r^2_{35.4}}},$$

$$r_{23.45} = \frac{r_{23.4} - r_{25.4}\,r_{35.4}}{\sqrt{1 - r^2_{25.4}}\sqrt{1 - r^2_{35.4}}}.$$

To compute these 3 r's we need 6 r's of the first order, viz.,—

$r_{12.4}$	$r_{15.4}$	$r_{13.4}$
$r_{25.4}$	$r_{23.4}$	$r_{35.4}$

(*b*)

To find: $r_{13.245}$ write as $r_{13.452}$. Then—

$$r_{13.452} = \frac{r_{13.45} - r_{12.45}\,r_{23.45}}{\sqrt{1 - r^2_{12.45}}\sqrt{1 - r^2_{23.45}}}.$$

To compute this r we need no partial r's other than those already found in (*a*).

(*c*)

To find: $r_{14.235}$ write without change—

$$r_{14.235}=\frac{r_{14.23}-r_{15.23}\,r_{45.23}}{\sqrt{1-r^2_{15.23}}\sqrt{1-r^2_{45.23}}}.$$

To compute this r we need 3 partial r's of the second order, viz.,—

$$r_{14.23}=\frac{r_{14.2}-r_{13.2}\,r_{34.2}}{\sqrt{1-r^2_{13.2}}\sqrt{1-r^2_{34.2}}}.$$

$$r_{15.23}=\frac{r_{15.2}-r_{13.2}\,r_{35.2}}{\sqrt{1-r^2_{13.2}}\sqrt{1-r^2_{35.2}}}.$$

$$r_{45.23}=\frac{r_{45.2}-r_{34.2}\,r_{35.2}}{\sqrt{1-r^2_{34.2}}\sqrt{1-r^2_{35.2}}}.$$

To compute these r's we need 6 r's of the first order, viz.,—

$r_{14\ 2}$	$r_{13.2}$	$r_{15.2}$
$r_{34.2}$	$r_{35.2}$	$r_{45.2}$

(*d*)

To find: $r_{15.234}$ write without change—

$$r_{15.234}=\frac{r_{15.23}-r_{14.23}\,r_{45.23}}{\sqrt{1-r^2_{14.23}}\sqrt{1-r^2_{45.23}}}.$$

To compute this r we need no partials other than those already found in (*c*).

[Note that we must compute a minimum of 4 third order r's, 6 second order r's, and 12 first order r's, 22 in all.]

(4) Standard Deviations. The 5 fourth order σ's required may be found from the following forms which make use of only those partial r's already computed in **(3)**:

$$\sigma_{1.2345}=\sigma_1\sqrt{1-r^2_{12}}\sqrt{1-r^2_{13\ 2}}\sqrt{1-r^2_{14.23}}\sqrt{1-r^2_{15.234}}\quad . \quad (50)$$

$$\sigma_{2.1345}\ (\text{i.e., } \sigma_{2.4531})=\sigma_2\sqrt{1-r^2_{24}}\sqrt{1-r^2_{25.4}}\sqrt{1-r^2_{23.45}}\sqrt{1-r^2_{12.345}}$$

$$\sigma_{3.1245}\ (\text{i.e., } \sigma_{3.4521})=\sigma_3\sqrt{1-r^2_{34}}\sqrt{1-r^2_{35.4}}\sqrt{1-r^2_{23.45}}\sqrt{1-r^2_{13.245}}$$

$$\sigma_{4.1235}\ (\text{i.e., } \sigma_{4.2351})=\sigma_4\sqrt{1-r^2_{24}}\sqrt{1-r^2_{34.2}}\sqrt{1-r^2_{45.23}}\sqrt{1-r^2_{14.235}}$$

$$\sigma_{5.1234}\ (\text{i.e., } \sigma_{5.2341})=\sigma_5\sqrt{1-r^2_{25}}\sqrt{1-r^2_{35.2}}\sqrt{1-r^2_{45.23}}\sqrt{1-r^2_{15.234}}$$

(5) Standard Error of Estimate $\sigma_{(\text{est. } X_1)}$

$$\sigma_{(\text{est. } X_1)}=\sigma_{1.2345}\ \text{[see (4) above for value]} \quad . \quad . \quad . \quad (54)$$

$$PE_{(\text{est. } X_1)}=.6745\ \sigma_{(\text{est. } X_1)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (55)$$

(6) **Coefficient of Multiple Correlation, *R*.**

$$R_{1(2345)} = \sqrt{1 - \frac{\sigma^2_{1.2345}}{\sigma^2_1}} \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (56)$$

which may be written also as

$$R_{1(2345)} = \sqrt{1 - [(1 - r^2_{12})(1 - r^2_{13.2})(1 - r^2_{14.23})(1 - r^2_{15.234})]},$$

and checked by

$$R_{1(2345)} = \sqrt{1 - [(1 - r^2_{14})(1 - r^2_{15.4})(1 - r^2_{12.45})(1 - r^2_{13.245})]}.$$

IV. A Multiple Correlation Problem with Four Variables

In Section II we found that a student's honor points (X_1) could be estimated with a considerable degree of accuracy from a knowledge of his general intelligence score (X_2) and the number of hours he spends in study per week (X_3). The $PE_{(\text{est. } X_1)}$ made in estimating individual scores from this three-variable regression equation was found to be 4.28 points; and the coefficient of multiple correlation, $R_{1(23)}$ which indicates, in general, how well the estimated scores represent the actual scores was .824. Now suppose that we add to the two independent variables X_2 and X_3 a third factor X_4—e.g., the quality of the preparatory work done by the student in High School.[1] This will give us *three* independent variables from which to estimate the dependent variable honor points, and the question arises:—with how much greater accuracy will this additional factor enable us to predict academic success?

The answer to this question will be found in Table XXVIII, which gives a complete solution of this problem, following the scheme outlined for four-variable problems in Section III(6). Some additional discussion of procedure and methods and several points to be especially noted are given in the following paragraphs.

Remember first of all that the mean and the σ of each set of measures must be known as well as their 6 intercorrelations,

[1] This was measured by the average grade obtained in the work offered for entrance to College. May, *Predicting Academic Success*, Journal of Educational Psychology, Vol. XIV, 434–436.

r's of the zero order. The calculation of these 6 intercorrelations is actually the most laborious part of the solution of a multiple correlation problem—in spite of the fact that we have passed it over with little comment heretofore—since a separate correlation table must be drawn up for each r.

(1) The discussion from here on[1] follows the outline given in (6) on page 240. Thus, before calculating any partial r's, we write the regression equation, and from it deduce what partial r's and higher order σ's will be required.

(2) It is clear from the regression coefficients that we shall need three partial r's of the second order:—viz., $r_{12.34}$, $r_{13.24}$, and $r_{14.23}$; and four partial σ's of the third order, viz., $\sigma_{1.234}$, $\sigma_{2.134}$, $\sigma_{3.124}$, and $\sigma_{4.123}$, in order to evaluate the constants in the regression equation. Only the partial r's actually required in the regression equation need be calculated.

(3) In order to find $r_{12.34}$ we shall need three first order partial r's, viz., $r_{12.3}$, $r_{14.3}$, and $r_{24.3}$; and to find $r_{13.24}$ we shall need, again, three first order partial r's, viz., $r_{13.2}$, $r_{14.2}$, and $r_{34.2}$. To find the last second order partial, $r_{14.23}$, no additional first order r's are required other than those already found. A minimum of 9 partial r's, therefore, is required in all.

The partial $r_{12.34}$ gives the net correlation between (1) honor points and (2) general intelligence when *both* (3) study hours and (4) average High School grades have been eliminated as variable factors or held constant. In like manner, $r_{13.24}$ gives the net correlation between (1) honor points and (3) study hours when both (2) general intelligence and (4) average High School grades are held constant. The first second order partial r, i.e., $r_{12.34}$, equals .764 and is but slightly reduced from $r_{12.3}$ which equals .802; while the second partial $r_{13.24} = .676$, and is also but slightly less than $r_{13.2}$ which equals .707. This comparison of partial r's shows the relatively small influence of High School grades on the net correlation between (1) honor points and (3) study hours with general intelligence constant, as well as the small influence of this factor on the net correlation

[1] See Table XXVIII. The divisions in the text parallel those in the table.

between (1) honor points and (2) general intelligence for study constant. Notice, however, that while the zero order coefficient of correlation between (1) honor points and (4) average High School grades, i.e., r_{14} is .40, $r_{14.2} = .246$, $r_{14.3} = .387$, and $r_{14.23} = .088$. Evidently, nearly all of the correlation which appears between (1) honor points and (4) average High School grades may be attributed to the common dependence of these two factors on (2) general intelligence and to a somewhat lesser degree on (3) study hours.

(4) By using the forms given in (**6**) page 240, we are enabled to calculate the four third order σ's required by the regression coefficients without the necessity of finding any additional partial r's (see page 234). These partial σ's viz., $\sigma_{1.234}$, $\sigma_{2.134}$, etc., give the *net* variability of the distribution of measures denoted by the *primary* subscripts when the influence of *all three* of the other factors (secondary subscripts) has been excluded. To take a single example, $\sigma_{1.234}$ is 6.31 as against a σ_1 of 11.2, which means, concretely, that if each of the 450 students in the group were exactly alike as regards (2) general intelligence, (3) study-hours, and (4) average High School grades, the σ of their distribution of honor points would be only about half as large as the observed σ:—the σ of the group in which these factors differ in weight or value.

The computation of the regression coefficients is simply a matter of combining the partial r's and σ's already found. When this has been done, we may substitute in the regression equation to find $x_1 = .55x_2 + 1.07x_3 + .083x_4$, or multiplying by 12.5 (a convenient constant), (the number of honor points) = 7 (score on general intelligence test) + 13 (the number of hours spent per week in study) + 1 (average High School grades). In Score Form the regression equation becomes $X_1 = .55X_2 + 1.07X_3 + .083X_4 - 69$.

It is clear from the regression equations that the number of hours spent in study has *twice* the weight of the score on general intelligence test and *thirteen* times the weight of the average High School grades, in determining the number of

honor points which a student will most probably receive at the end of the first semester. Apparently (as noted above), the average High School grades have relatively little influence on honor points as compared with the other factors in the equation.

(5) Still further evidence of the small importance of High School grades in improving the estimate of honor points is to be seen in the size of the $PE_{(\text{est. } X_1)}$. The PE of estimate made in predicting honor points from the present equation is 4.26 points as compared with a $PE_{(\text{est } X_1)}$ of 4.28 points made in using the regression equation which does not include High School grades (see page 230). This means that we can estimate the number of honor points which a student will receive, knowing his general intelligence score and the number of hours he spends in study per week, with but slightly greater error than when we know in addition to these two the average grade he has received in High School also. It would seem apparent, therefore, that the work required to build up a regression equation which will include the latter factor is hardly worth while.

(6) The multiple coefficient of correlation, $R_{1(234)}$ is .826 as compared with the $R_{1(23)}$ of .824. A comparison of these multiple coefficients further substantiates the conclusion that High School grades contribute practically nothing to the reliability of an honor point estimate.

It will be of considerable interest to compare the reliability of our estimate of honor points when the factors, singly and in combination, are taken into account. In this way the "prognostic" value of the multiple regression equation—as shown by the size of $\sigma_{(\text{est. } X_1)}$—will be more readily appreciated. The standard errors of estimate and the coefficients of correlation for the different factors taken singly and in combination are given below:

Dependent Variable: (Honor Points X_1)	$\sigma_{(\text{est. } X_1)}$	Coefficients of Correlation
$X_1 = .43X_2 - 24.76$	8.96	$r_{12} = .60$
$X_1 = .60X_3 + 4.1$	10.61	$r_{13} = .32$
$X_1 = .57X_2 + 1.13X_3 - 66$	6.34	$R_{1(23)} = .824$
$X_1 = .55X_2 + 1.07X_3 + .083X_4 - 69$	6.31	$R_{1(234)} = .826$

TABLE XXVIII

The Solution of a Correlation Problem Involving Four Variables

Computation of Means, σ's, and Intercorrelations

1. Honor Points	2. General Intelligence	3. Hours of Study per Week	4. Average High School Grade
$M_1 = 18.5$	$M_2 = 100.6$	$M_3 = 24$	$M_4 = 79$
$\sigma_1 = 11.2$	$\sigma_2 = 15.8$	$\sigma_3 = 6$	$\sigma_4 = 7.5$
$r_{12} = .60$	$r_{23} = -.35$	$r_{34} = .11$	
$r_{13} = .32$	$r_{24} = .36$		
$r_{14} = .40$			

For scheme of solution from here on, see page 240.

(1) Regression Equation:

$$X_1 = b_{12.34}X_2 + b_{13.24}X_3 + b_{14.23}X_4 + K \qquad (52)$$

(2) Regression Coefficients:

$$b_{12.34} = r_{12.34}\frac{\sigma_{1.234}}{\sigma_{2.134}} \qquad (53)$$

$$b_{13.24} = r_{13.24}\frac{\sigma_{1.234}}{\sigma_{3.124}}$$

$$b_{14.23} = r_{14.23}\frac{\sigma_{1.234}}{\sigma_{4.123}}$$

TABLE XXVIII—Continued

(3) Partial *r*'s. (49)

To find:

$$r_{12.34}=\frac{r_{12.3}-r_{14.3}r_{24.3}}{\sqrt{1-r^2_{14.3}}\sqrt{1-r^2_{24.3}}}$$

We must find 3 partial *r*'s of the first order:

$$r_{12.3}=\frac{r_{12}-r_{13}\cdot r_{23}}{\sqrt{1-r^2_{13}}\sqrt{1-r^2_{23}}}$$

$$=\frac{.60-.32\times-.35}{.9474\times.9367}=.802$$

$$r_{14.3}=\frac{r_{14}-r_{13}r_{34}}{\sqrt{1-r^2_{13}}\sqrt{1-r^2_{34}}}$$

$$=\frac{.40-.32\times.11}{.9474\times.9939}=.387$$

$$r_{24.3}=\frac{r_{24}-r_{23}r_{34}}{\sqrt{1-r^2_{23}}\sqrt{1-r^2_{34}}}$$

$$=\frac{.36-(-.35)\times.11}{.9367\times.9939}=.428$$

From the above:

$$r_{12.34}=\frac{.802-.387\times.428}{.9208\times.9028}=.764$$

To find:

$$r_{13.24}=\frac{r_{13.2}-r_{14.2}r_{34.2}}{\sqrt{1-r^2_{14.2}}\sqrt{1-r^2_{34.2}}}$$

We must find 3 partial *r*'s of the first order:

$$r_{13.2}=\frac{r_{13}-r_{12}r_{23}}{\sqrt{1-r^2_{12}}\sqrt{1-r^2_{23}}}$$

$$=\frac{.32-.60\times-.35}{.8\times.9367}=.707$$

$$r_{14.2}=\frac{r_{14}-r_{12}r_{24}}{\sqrt{1-r^2_{12}}\sqrt{1-r^2_{24}}}$$

$$=\frac{.40-.60\times.36}{.8\times.9330}=.246$$

$$r_{34.2}=\frac{r_{34}-r_{23}r_{24}}{\sqrt{1-r^2_{23}}\sqrt{1-r^2_{24}}}$$

$$=\frac{.11-(-.35)\times.36}{.9367\times.9330}=.270$$

From the above:

$$r_{13.24}=\frac{.707-.246\times.27}{.9682\times.9629}=.676$$

To find:

$$r_{14.23}=\frac{r_{14.2}-r_{13.2}r_{34.2}}{\sqrt{1-r^2_{13.2}}\sqrt{1-r^2_{34.2}}}$$

No first order partials required other than the 6 already found:

From the above:

$$r_{14.23}=\frac{.246-.707\times.27}{.7042\times.9629}=.088$$

TABLE XXVIII—Continued

(4) Standard Deviations: (50)

$$\sigma_{1.234}=\sigma_1\sqrt{1-r^2_{12}}\sqrt{1-r^2_{13.2}}\sqrt{1-r^2_{14.23}}=11.2\times.8\times.7072\times.9968 \quad =6.31$$

$$\sigma_{2.134}=\sigma_2\sqrt{1-r^2_{23}}\sqrt{1-r^2_{24.3}}\sqrt{1-r^2_{12.34}}=15.8\times.9367\times.9028\times.6499=8.68$$

$$\sigma_{3.124}=\sigma_3\sqrt{1-r^2_{23}}\sqrt{1-r^2_{34.2}}\sqrt{1-r^2_{13.24}}=6\times.9367\times.9629\times.7332 \quad =3.97$$

$$\sigma_{4.123}=\sigma_4\sqrt{1-r^2_{34}}\sqrt{1-r^2_{24.3}}\sqrt{1-r^2_{14.23}}=7.5\times.9939\times.9028\times.9959 \quad =6.71$$

The computation of partial regression coefficients $b_{12.34}$, $b_{13.24}$ and $b_{14.23}$.

$$b_{12.34}=r_{12.34}\frac{\sigma_{1.234}}{\sigma_{2.134}}=.764\times\frac{6.31}{8.68}=\ .55 \quad \ldots \quad (53)$$

$$b_{13.24}=r_{13.24}\frac{\sigma_{1.234}}{\sigma_{3.124}}=.676\times\frac{6.31}{3.97}=1.07$$

$$b_{14.23}=r_{14.23}\frac{\sigma_{1.234}}{\sigma_{4.123}}=.088\times\frac{6.31}{6.71}=\ .083$$

Substituting values of the b's, regression equation becomes: $x_1=.55x_2+1.07x_3+.083x_4$ (Deviation Form)

or since $(X_1-18.5)=.55(X_2-100.6)+1.07(X_3-24)+.083(X_4-79)$

$$X_1=.55X_2+1.07X_3+.083X_4-69 \quad \text{(Score Form)}$$

(5) The Standard Error of Estimate:

$$\sigma_{(\text{est. }X_1)}=\sigma_{1.234}=6.31 \quad \ldots \quad (54)$$

$$PE_{(\text{est. }X_1)}=.6745\times6.31=4.26 \quad \ldots \quad (55)$$

(6) Computation of R, coefficient of multiple correlation.

$$R_{1(234)}=\sqrt{1-\frac{\sigma^2_{1.234}}{\sigma^2_1}}=\sqrt{1-\frac{(6.31)^2}{(11.2)^2}}=.826 \quad \ldots \quad (56)$$

Check: $R=\sqrt{1-[(1-r^2_{13})(1-r^2_{14.3})(1-r^2_{12.34})]}=\sqrt{1-[.8976\times.8552\times.43]}=.826$

The important fact here is that $\sigma_{(\text{est. } X_1)}$ is considerably less, and the correlation considerably greater, when X_2 and X_3 are taken together than when either is taken alone. The standard error of estimate and the R improve very slightly when X_4 is added to X_2 and X_3. It is very probable that by an extension of the method of partial and multiple correlation to include other variables in addition to those we already have, the $\sigma_{(\text{est. } X_1)}$ of our problem could be still further reduced and R increased.

Before working out a regression equation containing added variable or variables the "predictive value" of the "new" equation should be found by computing $\sigma_{(\text{est. } X_1)}$ or R. This will enable us to determine what the effect will be of adding another variable or variables, and whether $\sigma_{(\text{est. } X_1)}$ is sufficiently reduced or R sufficiently increased to justify the additional calculation. In the present problem, for instance, either $\sigma_{(\text{est. } X_1)}$ or $R_{1(234)}$ would have told us that average High School grades add practically nothing to the predictive value of a regression equation which already contains the two variables general intelligence and number of hours spent on the average in study each week.

V. The Value and Use of Partial and Multiple Correlation

1. The Value and Use of Partial Correlation in Analysis and in Causal Investigations

Partial correlation is of considerable importance in the *analysis* of the part played by each of several factors in a total result, inasmuch as it enables us to find the *net* relationship between two sets of scores or measures when the influence of one or more other factors is excluded. A concrete illustration of this use of partial correlation may be cited from the work of Cyril Burt.[1] Burt wished to find how much a child's mental age—as given by the Binet tests—influenced his school attainment. His subjects were 300 children from 7 to 14 years old.

[1] Burt, Cyril, *Mental and Scholastic Tests*, London, 1921, pp. 180–184.

Each child's (1) *MA* (Binet) was found; likewise his (2) scholastic achievement as measured by educational examinations and checked by teachers; and (3) his chronological age. The "entire" coefficient of correlation between Binet *MA* and scholastic achievement (r_{12}) was .91. When chronological age (3) was held constant, the partial r ($r_{12.3}$) between Binet *MA* and scholastic achievement dropped to .68. This shows, in the first place, that age has a decided effect on the observed correlation between *MA* and school work—that it tends to increase or "dilate" the obtained r. This dilation is due to the fact that both *MA* and school attainment tend to increase with chronological age, and hence this common dependence on chronological age is sufficient to bring about a considerable "boost" in the observed correlation. In the second place, the $r_{12.3} = .68$ indicates that a substantial relation remains between *MA* and school work when age conditions are uniform. In other words, Binet *MA* (intelligence) is a substantial factor in a pupil's school attainment irrespective of his chronological age. To take the analysis a step further, Burt found that the correlation between school work (2) and chronological age (3) (r_{23}), was .87; and that when the effect of Binet *MA* was held constant, the partial r between school work and chronological age ($r_{23.1}$), was .49. The persistence of a fairly high relation between school work and chronological age when intelligence is eliminated offers confirmatory evidence, according to Burt, of the "undue influence of age upon school classification." In these illustrations it is clear that the calculation of the partial r's is the first step in an analysis of the factors which determine school attainment. By an extension of this same method the influence of other factors may be excluded and net relations secured.

From the analyses made through the elimination of factors by partial correlation, we are often enabled to determine existing "causal" relationships. Thus Phillips[1] in a study of the

[1] Phillips, Frank M., *Application of Partial Correlation to a Health Problem.* Reprint No. 867 from Public Health Reports, Sept., 1923.

causes contributing to absence on account of sickness among government employees over the period of a year found that the observed correlation between absence (i.e., number of persons absent) and mean temperature on the day of absence ($r_{at.}$) was $-.37$. When the four factors (1) relative humidity at 8 a.m. on the day of absence; (2) relative humidity at noon of the previous day; (3) inches of rainfall on the day of absence; and (4) per cent of possible sunshine on the day of absence were held constant, the net correlation ($r_{at.1234}$) remaining between absence and temperature was $-.39$, practically the same as the original correlation. Since this was the only r of any size (the other r's both entire and partial were negligible) the obvious conclusion seems to be that of the factors studied, temperature on the day of absence is the most important secondary or contributing cause of absence. (The sickness must be taken, of course, as the primary cause of absence.) Here and elsewhere let it be understood that partial correlation has absolutely nothing to say about "causes," as such. The conclusion as to which of two factors is the cause and which the effect is a matter of common sense analysis. In the illustration given, the distinction between cause and effect is obvious.

Another interesting example of the use of partial correlation in a causal investigation is found in the work of Reavis.[1] This investigator undertook to ferret out the causes of attendance and non-attendance in rural schools. Certain factors, (1) distance from school, (2) age-grade relation, (3) kind of work done by the pupils, (4) training, experience, etc., of teacher, (5) school equipment, and (6) kind of community were taken as having more or less effect on school attendance. When partial correlation was applied to the problem, it was found that the entire coefficient of correlation between attendance and distance, and attendance and kind of community, were the least reduced. The first was lowered from $-.45$ to $-.43$; and the second from .30 to .28. Of all the factors selected, therefore, these two seem

[1] Reavis, George, *Factors Controlling Attendance in Rural Schools.* Teachers College, Columbia University, 1920.

to have the most direct or independent influence on school attendance. As in the problem cited above, the distinction between cause and effect in this illustration is clear:—it is evident that distance from school and kind of community are the causes and not the effects of attendance or non-attendance.

2. The Value of the Regression Equation in Prediction and Analysis

The value of the regression equation is twofold:[1] (1) In its usual form, it gives the *weights* to be assigned each of several independent variables, in order that X_1 (the dependent variable) may be predicted or forecasted with minimum error (see page 237). (2) In its "special" form it may be used to analyze—within certain limits—a given capacity or ability. We shall consider these two uses of the regression equation in order.

(1) It has already been stated that the regression equation enables us to combine two or more tests or other measures (independent variables, X_2, X_3, $\ldots X_n$) into a single value (X_1) in such a way as to give the best possible estimate of X_1. In the three-variable problem on page 228, for example, the regression equation gives us the best possible forecast of the number of honor points (X_1) which a student will receive, when we know his general intelligence score (X_2) and the average number of hours he spends per week in study (X_3). Moreover, once calculated, the regression equation may be used subsequently to estimate other student's scores in X_1 when only their scores in X_2 and X_3 are known. The value of the regression equation as a forecasting instrument is determined by the size of the standard error of estimate, and by the multiple coefficient of correlation.

A good illustration of the value of the regression equation in forecasting—taken from another field than psychology—is to be found in the work of Moore in forecasting the cotton crop in

[1] Kelley, T. L., *Tables to Facilitate the Calculation of Partial Coefficients of Correlation and Regression Equations*, Bulletin of the University of Texas, 1916, 27, p. 7.

the Southern States.[1] Taking the cotton crop in Georgia as the dependent variable (to cite a single example) and the May rainfall, June temperature, and August temperature as independent variables, Moore built up a regression equation from which it was possible to get a better forecast of the crop at the end of August than the official method of the U. S. Department of Agriculture could obtain from the condition of the crop in September. (By better forecast is meant a smaller error of prediction.)

In addition to its use as a forecasting instrument, the regression equation may be used also to determine the value or "weight" which each test in a battery should have in order that the composite scores obtained from the battery (group of tests) shall be the *best* possible estimates of that capacity which the whole battery of tests presumably measures. This is essentially the same problem as that of prediction or forecasting discussed in the last paragraph. Suppose, by way of illustration, that the problem is to devise a group test for measuring general intelligence; and that this battery is to consist of four tests. The first step is to secure some good "criterion"[2] of general intelligence. This may be (1) school grades, (2) teachers' estimates, (3) (1) and (2) combined, or (4) some standard intelligence examination, as for example, Stanford-Binet or Army Alpha. The next step is to select four tests which will separately give (1) high correlations with the criterion, and (2) low correlations with each other.[3] These two conditions guarantee that each test will measure *some aspect* or *phase* of the criterion; and further that each test will probably measure a different, or slightly different, phase of the criterion, since the low intercorrelations will prevent much duplication. Let us call the criterion X_c and the four tests of the battery X_1, X_2, X_3, and X_4. The regression equation in Score Form is

[1] Moore, H. L., *Forecasting the Yield and Price of Cotton*, 1917, pp. 108–115.
[2] See page 266 for definition of "criterion."
[3] The ideal battery of tests would consist of tests which correlate as high as possible with the criterion, and as low as possible with each other.

$X_c = AX_1 + BX_2 + CX_3 + DX_4 + K$: in which A, B, C, D, the regression coefficients, are the "*weights*" to be given the scores made on the four tests, and K is a numerical constant. Now to take a very simple case, suppose that $A=1$; $B=2$; $C=3$; and $D=4$. The regression equation then becomes $X_c = 1X_1 + 2X_2 + 3X_3 + 4X_4 + K$: which means that a subject's score on test No. 1 must be multiplied by 1, his score on test No. 2 by 2, his score on test No. 3 by 3, and his score on test No. 4 by 4 in order that his composite score on the battery may give the "best" estimate of his score on X_c, the criterion.

The regression equation may be said to furnish the *ideal* method of combining several tests into a team, since each test in a regression equation is weighted according to its correlation with the criterion, independently of the other tests in the team or battery. Under these conditions the standard error of estimate is a minimum while the correlation of the predicted X_c values and the actual X_c values (multiple R) is the *maximum* obtainable with the given set of tests. R tells the extent to which our team *represents* the criterion.

(2) The only difference between the usual or "regular" form of the regression equation and the "special" form to be considered now is that in the special form, the σ's of *all* of the different tests (or other measures) are taken as *equal.* This procedure eliminates differences in the size of the test units as well as differences in "spread" or variability, and enables us to determine (from the correlation alone) the relative weight with which each independent factor "enters into" or contributes to the dependent variable (the criterion) independently of the other factors. In this way, an analysis can be made of the importance of several different factors in some final result. *It is very important to remember, however, that in its special form, the regression equation cannot be used for forecasting.*

We may illustrate the special use of the regression equation with data taken from the three-variable problem on page 228. If X_1, honor points, be taken as the criterion, while X_2, general intelligence, and X_3, average number of hours spent in study

per week are, as before, the independent variables, the usual or "regular" regression equation is written:

$$X_1 = b_{12.3}X_2 + b_{13.2}X_3 + K.$$

Replacing the b's in this equation by means of formula (53),

$$X_1 = r_{12.3}\frac{\sigma_{1.23}}{\sigma_{2.13}}X_2 + r_{13.2}\frac{\sigma_{1.23}}{\sigma_{3.12}}X_3 + K;$$

and replacing the partial σ's [by formula (50)], we have

$$X_1 = r_{12.3}\,\frac{\sigma_1\sqrt{1-r^2_{13}}\sqrt{1-r^2_{12.3}}}{\sigma_2\sqrt{1-r^2_{23}}\sqrt{1-r^2_{12.3}}}\,X_2$$

$$+ r_{13.2}\,\frac{\sigma_1\sqrt{1-r^2_{12}}\sqrt{1-r^2_{13.2}}}{\sigma_3\sqrt{1-r^2_{23}}\sqrt{1-r^2_{13.2}}}\,X_3 + K.$$

Substituting numerical values for the r's and putting $\sigma_1 = \sigma_2 = \sigma_3$, we have

$$X_1 = .802\frac{(.9474)}{(.9368)}X_2 + .707\frac{(.8)}{(.9367)}X_3 + K,$$

or

$$X_1 = .8X_2 + .6X_3 + K.$$

This result may be interpreted to mean that in so far as the two factors, general intelligence and number of hours spent on the average in study per week, "enter into" the ability to get honor points, they contribute with the relative weight of .8 : .6 or 4 : 3. It must be clearly understood that this ratio refers to the relative contribution of the two factors *themselves* to the final result and *not* to the relative weights of their *scores*. The weight to be assigned each *score* is found from the regular regression equation given on page 229. It is of considerable interest, however, to note that while the scores on the general intelligence test and number of study hours are as 1 : 2, the actual contribution of these two factors to honor points (allowing for differences in units, variability, etc.) is as 4 : 3. Intelligence, therefore, as we should expect, has more weight than hours spent in study in determining the hypothetical ability

which we have called "academic success." Much of the weight which study-hours has is due to its relatively high negative correlation (−.35) with intelligence.

In concluding this discussion of partial and multiple correlation, certain limitations to the use of the method should be pointed out. In the first place, in order that partial coefficients of correlation be valid, it is necessary that all of the zero order coefficients be computed from data in which the regression is linear. Before calculating any partial r's, we should make sure that all zero order r's have linear regression: if there is any doubt as to linearity, the tests given on page 209 should be employed. In the second place, the number of cases must be large, especially if there are a number of variables, otherwise partial and multiple coefficients will have little significance. Coefficients which are misleadingly high may be obtained when studies which involve many variables are based upon relatively few cases. When the limitations and conditions mentioned are fully recognized and met, however, partial and multiple correlation furnishes us with an exact and powerful instrument for the analysis of problems which arise in mental and social measurements.

VI. Spurious Correlation[1]

The correlation between two sets of test scores is said to be "spurious" when it is due in whole or part to factors other than those which determine performance in the tests themselves. In general, the cause of spurious correlation may be said to lie in a failure to control conditions; and the most usual effect of this lack of control is a "boosting" or dilation of the coefficient. Some of the more general situations which may lead to spurious correlation are given under the following heads:

1. Spurious Correlation Due to the Heterogeneity of Material

We have already found occasion to show elsewhere (page 221) how a lack of uniformity in age conditions will lead to

[1] See also Chap. IV, p. 211.

correlation which is too high, i.e., is spurious. Differences in age within the group will lead to a distinctly higher correlation between two tests—when the test scores increase with age—than the correlation which we should obtain in a single age (a homogeneous) group. To cite a simple case, in a group of boys from 10 to 18 years old, a substantial correlation will appear between strength of grip and length of forearm, quite apart from any real relation, due solely to the fact that both of these physical attributes increase with age.

Failure to take account of the age factor is a prolific source of error in correlational work. In stating the correlation between two tests, or the reliability coefficient of a test, we should always be careful to specify the range of ages, grades, etc., in order to show the heterogeneity of the group. Without this information an r *per se* is practically valueless.

Many other factors besides age may lead to spurious correlation. To cite a familiar example:[1] if alcoholism, degeneracy and bad heredity are all positively related, the r between alcoholism and degeneracy will be too high (due to the indirect effect of heredity on both factors) unless the heredity influences are kept constant. Again, to take another example, suppose that we have found the scores on a general intelligence examination and a cancellation test for two distinctly different groups, e.g., 500 college seniors and 500 day laborers; and that the average ability in both tests is definitely higher in the college group. Now if the correlation between these tests is zero in each group *taken separately*, when the two groups are *combined* a positive correlation will be obtained due simply to the heterogeneity of the composite group.[2] Such a correlation is, of course, spurious.

To be valid, it is clear that a correlation must be freed of extraneous influences which affect the homogeneity of the material. When such influences cannot be determined quan-

[1] Kelley, T. L., *Tables to Facilitate the Calculation of Partial Coefficients of Correlation and Regression Equations*, Bull. Univ. Texas, 1916, No. 27.

[2] Otis, A. S., *Statistical Method in Educational Measurement*, 1925, pp. 334–336.

titatively, this is far from an easy task. Provided, however, the factor or factors producing heterogeneity are measurable, their influence may usually be allowed for by the method of partial correlation.

2. Spurious Index Correlation

It can be shown [1] that three variables X_1, X_2, and X_3 may be totally uncorrelated, and still a correlation between $Z_1 = \frac{X_1}{X_3}$, and $Z_2 = \frac{X_2}{X_3}$ may be obtained which is as large as .50. To take a concrete case, if two individuals observe a series of magnitudes (e.g., Galton Bar settings) independently, the absolute errors of observation (X_1 and X_2) may be uncorrelated, and still a distinct correlation appear between the errors made by the two observers when these are expressed as *per cents* of the magnitude observed (X_3). The spurious element here is, of course, the common factor, X_3, in the denominator of the ratios.

One of the commonest examples of spurious index correlation in psychology is found in the correlation of IQ's obtained from two different intelligence tests. If the IQ's of 500 children ranging in age from 3 to 14 years are calculated from two tests X_1 and X_2, the correlation between IQ_{x_1} and IQ_{x_2} will be considerably increased because of the presence of the common factor of chronological age X_3 $\left(\text{since } IQ = \frac{MA}{CA}\right)$ in the two series. The spurious element here may be eliminated by holding constant the common factor of age through partial correlation.

3. Spurious Correlation of a Single Test With a Composite of Which it is a Member

If the scores of several tests, X_1, X_2, X_3, etc., are averaged or added, and the composite scores, $X_{\text{com.}}$ correlated with the scores of any single test X_1, the correlation resulting will be too high (spurious) because of the presence of X_1 in the composite.

[1] Yule G. U., *An Introduction to the Theory of Statistics*, pp. 215–216.

The amount or degree of the spurious element is measured by the ratio $\frac{t}{s}$ in which t = the number of elements in the single test, and s = the number of elements in the composite[1] (see page 293). To illustrate: there are 20 items in the Number Series Completion Test of the Army Alpha, and 212 items in the whole test. Now if there were *no* correlation at all between the scores on Alpha and Completion there would still be a spurious correlation between the two tests equal to the ratio of the number of items in Completion to the total number of items in Alpha, i.e., $\frac{20}{212}$ or .094. A correlation obtained between Completion and Alpha, therefore, will be too high, due simply to the inclusion of the Completion items in both sets of data.

It should be noted that when several tests are all of the same—or approximately the same—length, the amount of spurious correlation which will result from correlating any single test with a composite of them all is approximately constant $\left(\frac{t}{s} \text{ is same}\right)$. For this reason it is valid to compare the correlations of the separate tests with the composite in order to discover which tests are most representative of the capacity measured by them all (see page 267).

VII. Summary of Formulas in Chapter V

1. Partial r's,

$$r_{12.34 \ldots n} = \frac{r_{12.34 \ldots (n-1)} - r_{1n.34 \ldots (n-1)} r_{2n.34 \ldots (n-1)}}{\sqrt{1 - r^2_{1n.34 \ldots (n-1)}} \sqrt{1 - r^2_{2n.34 \ldots (n-1)}}}. \quad . \quad (49)$$

2. Partial σ's,

$$\sigma_{1.234 \ldots n} = \sigma_1 \sqrt{1 - r^2_{12}} \sqrt{1 - r^2_{13.2}} \sqrt{1 - r^2_{14.23}} \ldots \sqrt{1 - r^2_{1n.23 \ldots (n-1)}}. \quad (50)$$

3. Regression Equation, Deviation Form,

$$x_1 = b_{12.3 \ldots n} x_2 + b_{13.2 \ldots n} x_3 \ldots + b_{1n.23 \ldots (n-1)} x_n. \quad (51)$$

[1] Musselman, J. R., *Spurious Correlation Applied to Urn Schemata*, Journal of American Statistical Association, Vol. XVIII, Sept., 1923.

4. Regression Equation, Score Form,

$$X_1 = b_{12.34\ldots n}X_2 + b_{13.24\ldots n}X_3 \ldots + b_{1n.23\ldots(n-1)}X_n + K. \quad (52)$$

5. Regression Coefficients,

$$b_{12.34\ldots n} = r_{12.34\ldots n}\frac{\sigma_{1.234\ldots n}}{\sigma_{2.134\ldots n}}. \quad (53)$$

6. Standard Error of Estimate,

$$\sigma_{(\text{est. } X_1)} = \sigma_{1.234\ldots n}. \quad (54)$$

7. Probable Error of Estimate,

$$PE_{(\text{est. } X_1)} = .6745 \times \sigma_{(\text{est. } X_1)}. \quad (55)$$

8. Multiple Coefficient of Correlation,

$$R_{1(23\ldots n)} = \sqrt{1 - \frac{\sigma^2_{1.23\ldots n}}{\sigma^2_1}}. \quad (56)$$

9. Formula for "Chance" R,

$$R = \frac{\sqrt{n-1}}{\sqrt{N}}. \quad (57)$$

10. Alternate formula for R,

$$R_{1(234\ldots n)} = \sqrt{1 - [(1 - r^2_{12})(1 - r^2_{13.2})\ldots(1 - r^2_{1n.23\ldots(n-1)})]}. \quad (58)$$

PROBLEMS

1. The r for intelligence and school achievement in a group of children 8 to 14 years old is .80. The r for intelligence and age in the same group is .70. The r for school achievement and age is .60. What will be the correlation between intelligence and school achievement in children of the same age?
2. The correlation between (1) Army Alpha and (2) Cancellation in a group of 100 freshmen is .20. The correlation between (1) Army Alpha and (3) Controlled Association in the same group is .70. The correlation between (2) Cancellation and (3) Controlled Association is .45. What is the net correlation between Alpha and Cancellation in this group? Between Alpha and Controlled Association? How do you interpret your results?

3. Given the following data:[1]

X_1=high school grade in mathematics.
X_2=grade in an English interest test.
X_3=grade in a history interest test.
X_4=grade in a mathematics interest test.

$\sigma_1=4.93$	$r_{12}=.20$	$r_{23}=.63$
$\sigma_2=3.13$	$r_{13}=.15$	$r_{24}=.21$
$\sigma_3=6.12$	$r_{14}=.24$	$r_{34}=.54$
$\sigma_4=4.64$		

(*a*) Work out the regression equation of X_1 on X_2, X_3, X_4.
(*b*) What are the relative weights of the three tests, X_2, X_3, and X_4, in determining the score on X_1?

4. The following records were secured from 450 Liberal Arts freshmen at Syracuse University:[2]

1. Honor points	2. Intell.	3. Aver. H. S. Grades	4. Units	5. Hours per week of study
$M_1=18.5$	$M_2=100.6$	$M_3=79$	$M_4=16.1$	$M_5=24$
$\sigma_1=11.2$	$\sigma_2=15.8$	$\sigma_3=7.5$	$\sigma_4=1.5$	$\sigma_5=6$

$r_{12}=.60$	$r_{23}=.36$	$r_{34}=.40$	$r_{45}=.25$
$r_{13}=.40$	$r_{24}=.20$	$r_{35}=.11$	
$r_{14}=.22$	$r_{25}=-.35$		
$r_{15}=.32$			

(*a*) Work out a regression equation with (1) honor points as the dependent variable.
(*b*) If a student has an intelligence score of 110, a High School average of 75, offers 15 units for entrance, and studies on the average 25 hours per week, what is his most probable number of honor points?

5. Using as much of the data in Example (4) as is necessary, find how many hours a student must study if he has an intelligence score of 120, and wants to make 20 honor points? (Hint: work

[1] Kelley, T. L., *Educational Guidance*, Teachers College, Contributions to Education, 1914, 71, p. 104.
[2] May, Mark A., *Predicting Academic Success*, Journal of Educational Psychology, 1923, Vol. XIV, 7, 429–440.

out the regression equation of study hours on honor points and intelligence and substitute the given values in the equation.)

6. Let X_1 be a criterion, and X_2 and X_3 two other tests. Correlations and σ's are as follows:

$r_{12} = .60$	$r_{23} = .20$	$\sigma_1 = 5.00$
$r_{13} = .50$		$\sigma_2 = 10.00$
		$\sigma_3 = 8.00$

How much more accurately can X_1 be predicted from X_2 and X_3 in combination than from either alone?

7. Given a team of two tests, each of which correlates .50 with a criterion. If the correlation of the two tests is .20,
   (*a*) How much would the addition of another test which correlates .50 with the criterion and .20 with each of the other tests improve the predictive value of the team?
   (*b*) How much would the addition of two such tests improve the predictive value of the team?
8. Two absolutely independent measures *B* and *C* completely determine a third measure *A*. If *B* correlates .50 with *A*, what is the correlation of *C* and *A*?
9. Using the data given in Example (1) above, analyze school achievement in terms of intelligence and age. What is the relative importance of the contribution made by these factors?
10. A group test contains 10 tests with a total of 200 items. One of the tests correlates .60 with the composite scores on the battery. If this test contains 15 items, how much of the given correlation is spurious?

Answers

1. $r = .67$.
2. The r between Alpha and Cancellation is $-.18$; between Alpha and Controlled Association, .70.
3. (*a*) $x_1 = .37x_2 - .11x_3 + .28x_4$.
   (*b*) Grade in mathematics $= 6.5$ (grade in English interest test) -2 (grade in history interest test) $+5$ (grade in mathematics interest test).

4. (*a*) $X_1 = .58X_2 + .14X_3 - 1.03X_4 + 1.10X_5 - 62$
   (*b*) 24 with a $PE_{(\text{est. } X_1)}$ of 4 points.
5. 18 hours with a $PE_{(\text{est. } X_1)}$ of 2.7 hours: 18 ± 2.7
6. From X_2 alone $\sigma_{(\text{est. } X_1)} = 4.0$
   From X_3 alone $\sigma_{(\text{est. } X_1)} = 4.3$
   From X_2 and X_3 $\sigma_{(\text{est. } X_1)} = 3.5$
7. (*a*) R increases from .64 to .73.
   (*b*) R increases from .64 to .79.
8. $r_{AC} = .866$.
9. Intelligence and age contribute in the ratio (approximately) of 10 : 1.
10. .075.

CHAPTER VI

SOME APPLICATIONS OF STATISTICAL METHOD AND TECHNIQUE TO TESTS AND TEST RESULTS

To treat properly all of the statistical methods which may be applied to tests would require not a single chapter but a volume in itself. The aim of the present chapter, therefore, is to consider simply those methods—having to do largely with correlation and reliability—which are deemed essential (1) in the treatment of ordinary problems involving tests and (2) as a foundation for more advanced work in methods of treating test results.

I. The Validity of Test Scores

The validity of any measuring instrument depends on the *fidelity* with which it measures whatever it purports to measure. A yardstick is "valid" when measurements made by it can be checked by other measuring instruments. And in like manner a test is valid when the capacity which it measures corresponds to the same capacity as otherwise objectively measured and defined.

1. Validity Determined through Correlation with a Criterion

The validity of a test is usually determined by finding the correlation between the test and some independent criterion. A criterion is defined as that measure in terms of which the value of a test is estimated or judged. The criterion of a general intelligence test, for example, may be school marks, or ratings for intelligence, or some other test believed to be valid.[1]

[1] Stanford-Binet is often taken as a reliable criterion of general intelligence. For example, see Herring Revision of Binet-Simon tests.

The criterion for a trade test is actual ability in the trade. A high correlation between a test and its criterion may be taken as evidence of validity, provided both the test and the criterion are reliable. Before accepting criterion-correlations as final, however, we must know the reliability of our test, and if possible, we should know also the reliability of our criterion.[1]

2. Indirect Measures of Validity

When a reliable criterion is not available, indirect methods must be employed to determine validity. One indirect method is to combine the scores on a number of tests of the same general function and to judge as best (most valid for the function) that test which correlates highest with the average of all. Thus Whitley[2] found for three discrimination tests, Naming Colors, Naming Forms, and Naming Objects, the following correlations:[3]

$$\text{Average of all three tests with} \begin{cases} \text{Naming Colors} & r = .67 \\ \text{Naming Forms} & r = .99 \\ \text{Naming Objects} & r = .96 \end{cases}$$

She concludes that "Naming Forms seems more a typical test in so far as it measures an ability common to these three tests." In the absence of an independent measure of the function the average of several tests of that function may be taken as one criterion.

A second indirect method of measuring validity is to find correlations between the given test and other tests, in this way discovering some of the facts which the test does, and does not, measure. For example, tests of Controlled Association, e.g., Opposites, Logical Relations, etc., correlate much more highly with tests of general intelligence and "reasoning" than with tests of Cancellation or Color-Naming. The first group of tests is, therefore, a better (more valid) measure of the capacity

[1] Kelley, T. L., *The Reliability of Test Scores,* Journal Educational Research, 1921, Vol. 3, 5, p. 370.

[2] *Tests for Individual Differences*, Archives of Psychology, 1911, 19, p. 78.

[3] The "spurious" element here is constant provided the tests are all of practically the same length (see page 261).

measured by the general intelligence and reasoning tests than the second group. (Indirect measures of this sort are advisable only in the absence of more direct and valid criteria.)

The absence of valid criteria for many of his tests forces the careful psychologist to define tests strictly in terms of what they actually do. Hence the tendency of present-day testers is to call a test by some descriptive name rather than in terms of some more or less well-defined "mental function." Accordingly, we have Opposites Tests, and Completion Tests rather than tests of Association or Reasoning.

II. The Reliability of Test Scores

1. The Reliability of a Test as Measured by Its Self-Correlation

A. The "Reliability Coefficient"

The reliability of a test (or of any measuring instrument) is determined by the *consistency* with which it measures the capacity of those taking it. If a group repeats a test and each individual in the group scores close to his first record, we regard the test as reliable. If, however, there are large positive and negative differences between the scores made by individuals on the first and second giving of the test over and above the practice effect[1]—and if such differences occur in a large number of cases—obviously the test is inconsistent and unreliable. One method of measuring the reliability of a test is to correlate the scores made on the test by a given group with the scores made on the same or a duplicate test by the same group. This is the method of self-correlation; and the r so found is called the "reliability coefficient."

When the reliability coefficient of a test is 1.00, the test is an absolutely accurate measure of whatever capacity it tests, and when the reliability coefficient is .00 the test has just no reliability. The lower the reliability coefficient the less the reliability or consistency of the test as a measuring instrument.

[1] Practice, since it serves to increase all scores proportionally, does not affect self-correlation. It does, however, introduce a constant error.

How high should self-correlation be in order to indicate a satisfactory reliability? This is an important question and its answer depends largely on the nature of the test and the size and variability of the group for whom the test is intended. Most makers of general intelligence tests demand a reliability coefficient of at least .90 between duplicate forms of their tests for unselected groups of the same chronological age. To be a reliable measure of capacity, a mental or physical test should—generally speaking—have a minimum reliability coefficient of at least .80. This minimum will vary with the group, however, as the reliability coefficient is considerably affected by the range of scores made on the test (see page 271). For this reason, in giving the reliability coefficient of a test the size and variability of the group measured should always be stated.

B. Effect on Reliability of Lengthening or Repeating the Test

If the self-correlation of a test is unsatisfactory two courses are open: (1) we can lengthen the test until the reliability is greater; or (2) we can repeat the test and its duplicate *twice* each, average the two series of scores, and correlate these averages. If after (2) the reliability coefficient is still too low, we can repeat the test and its duplicate, three, four, or as many times as is necessary to secure the desired reliability coefficient. To do either (1) or (2) empirically would require a considerable amount of time and labor; hence it is fortunate that a good measure of the effect of (1) or (2) may be expeditiously secured by applying Spearman's (sometimes called Brown's[1]) "prophecy" formula:

$$r_x = \frac{Nr}{1+(N-1)r}. \quad \quad . \ . \ . \ . \ . \ . \ . \quad (59)$$

To illustrate the application of this formula, suppose (*a*) that the self-correlation of a test is .70 and that we wish to know what will be the effect of *doubling* the length of the test

[1] Brown, Wm., *The Essentials of Mental Measurement*, 1911, p. 102.

on its reliability. Substituting $r = .70$ and $N = 2$ in the formula, and solving for r_x we have

$$r_x = \frac{2 \times .70}{1 + .70} = .82.$$

Doubling the test's length, therefore, increases the self-correlation from .70 to .82. Instead of doubling the length of the test, we may give it and its duplicate *twice* each, average the two scores made by each individual in the two series, and correlate these averages. The result will be the same (as far as purely statistical factors are concerned) as that obtained by doubling the length of the test.

The "prophecy" formula may be used in another way. Suppose (*b*) that the self-correlation of a test or the correlation of the test and its duplicate is .80. How much will the test have to be lengthened (or how many times repeated) in order to insure a reliability coefficient (r_x) of .95? Substituting $r = .80$ and $r_x = .95$ in the formula, and solving for N,—

$$.95 = \frac{.8N}{1 + .8N - .8} = \frac{.8N}{.2 + .8N},$$

$$.04N = .19$$

$$N = 4.75 \text{ or } 5.00 \text{ (in whole numbers).}$$

The test must be 5 times its present length or repeated (together with its duplicate) 5 times in order to raise the self-correlation from .80 to .95.

When a test is increased in length, e.g., doubled or tripled, the items or questions added must always be equal in reliability to the reliability of the original test, if the results from the prophecy formula are to be valid. Provided this condition is satisfied, it is evident that if we increased the length of a test indefinitely we could—theoretically—raise its self-correlation to any desired figure. This seems scarcely reasonable, however; and there is evidence to indicate that while the reliability

coefficient increases according to the formula for the first four or five pooled tests, thereafter it increases "more slowly than the prediction formula would lead us to expect." [1]

C. Coefficient of Reliability from One Application of a Test

If a test has no duplicate and cannot well be repeated, we may measure the reliability of half of the test and then by Spearman's formula find the reliability of the whole test. The procedure is as follows: First, we make up two independent sets of scores by combining, say, alternate exercises in the test. For example, one set of scores may be the performance on the odd exercises, e.g., 1, 3, 5, etc.; the other set the performance on the even exercises, e.g., 2, 4, 6, etc.; or some other plan may be used.[2] These two sets of scores are now correlated to find the reliability coefficient of the *half* test. If the self-correlation of the half test so found is called r_h, substituting $N=2$ in Spearman's formula, we can calculate the reliability of the whole test by the formula,

$$r_x=\frac{2r_h}{1+r_h}. \qquad (60)$$

In using this formula we make the assumption that the halves of the test as we have made them up are approximately equivalent in difficulty and content.

D. Dependence of the Reliability Coefficient on the Size and Variability of the Group

The coefficient of reliability obtained from a test and its duplicate given to the pupils of a single grade cannot be taken as indicative of the same degree of reliability as the identical coefficient obtained from a group composed of pupils spread over several grades. This is due to the fact that the heterogeneity—

[1] Holzinger, Karl J., *Note on the Use of Spearman's Prophecy Formula for Reliability*, Journal Educational Psychology, 1923, Vol. XIV, 5, pp. 301-305.

[2] Ruch, G. M., and Del Manzo, M. C., *The Downey Will Temperament Group Test; Analysis of its Reliability and Validity*, Journal Applied Psychology, Vol. VII, 1, 1923, p. 65.

the size, and spread—of the two groups is different. Recently Kelley [1] has devised a formula from which, knowing the reliability coefficient of a test, say, in a group composed of pupils from a single grade, we can determine what the reliability coefficient of the same test must be in a group composed of pupils from several grades in order that the test be equally effective in both ranges. The formula is

$$\frac{\sigma}{\Sigma}=\frac{\sqrt{1-R}}{\sqrt{1-r}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (61)$$

in which σ and Σ are the σ's of the scores in the *small* and *large* groups, respectively, and r and R are the reliability coefficients of the test in the small and large groups. To illustrate, suppose that in a single grade $r=.50$ and $\sigma=5.00$; and that in a large group made up of children from grades 3 to 8, inclusive, $\Sigma=15$. What R (i.e., reliability coefficient) must the test yield in the large group in order to be as effective here as in the small group? Substituting for σ, Σ, and r in the formula, $R=.94$,—which means that a reliability coefficient of .50 in the small group indicates the *same* degree of reliability as a reliability coefficient of .94 in the group in which the range of "talent" is three times as great.

This formula may be used to determine whether a test is equally effective in parts of the range (σ) as in the whole range (Σ); or in one range as in another. It also serves to make clear the necessity of always giving the size and spread of the group in stating and interpreting reliability coefficients.[2]

2. The Index of Reliability

By an individual's "true" score in a test is meant the average of a very large number of measurements made of the given individual on the same or duplicate tests under precisely

[1] *The Reliability of Test Scores,* Journal Educational Research, 1921, Vol. III, 5, pp. 370–379.

[2] Otis, A. S., *Statistical Method in Educational Measurement,* 1925, pp. 253–254.

the same conditions. It has been shown [1] that the correlation between a series of obtained scores and their corresponding "true" scores may be found from the formula

$$r_{\text{obt. true}} = \sqrt{r_{12}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (62)$$

in which r_{12} is the self-correlation or the reliability coefficient obtained from duplicate forms of the test. Given the reliability coefficient, therefore, it is possible to secure the coefficient of correlation between a set of obtained scores and their corresponding true scores. This coefficient, $r_{\text{obt. true}}$, is called the "index of reliability," and is the *maximum* value which the reliability coefficient, r_{12}, can take. This will be seen to follow from the fact that "the highest possible correlation which can be obtained (except as chance might occasionally lead to higher spurious correlation) between a test and a second measure is with that which truly represents what the test actually measures, —that is, the correlation between the test and the true scores of individuals in just such tests." [2] Since r_{12} is usually less than 1.00, $r_{\text{obt. true}}$ is nearly always greater than r_{12}.

To illustrate the index of reliability, suppose that for a given group, $r_{12} = .64$. Then $r_{\text{obt. true}} = \sqrt{.64}$ or .80, and .80 is the highest self-correlation which can be obtained (except by chance) with this test in its present form. The index of reliability is a useful and easily interpreted measure of a test's reliability, since by simply extracting the square root of an obtained reliability coefficient we can find the maximum reliability which the test is capable of yielding. Thus, if $r_{12} = .25$, so that $r_{\text{obt. true}} = \sqrt{.25}$ or .50, it is obviously a waste of time to continue using the test without lengthening or otherwise improving it.

[1] Kelley, T. L., *A Simplified Method of Using Scaled Data for Purposes of Testing*. School and Society, 1916, Vol. IV; 34, 71.

[2] Kelley, T. L., *The Reliability of Test Scores*, Journal of Educational Research, 1921, Vol. III, 5, 327.

3. The Standard Error and Probable Error of Measurement $\sigma_{(M)}$ and $PE_{(M)}$

We have seen that the reliability of a test may be measured in terms of (1) its reliability coefficient, and (2) its index of reliability. Still another way of measuring the reliability of a test is to determine how closely a score obtained on the given test approximates its corresponding true score. (True scores have been defined on page 272.) An obtained score will usually differ in some degree from its corresponding true score due to the presence of two sorts of errors,—constant errors and variable errors. Constant errors, since their weight is all in one direction, do not affect self-correlation, and can usually be ruled out or their influence measured. Variable errors, however, since they may be either positive or negative, are less easily eliminated than constant errors, and hence are more effective in producing departures of obtained scores from corresponding true scores.

The measurement of the influence of variable errors, therefore, becomes a matter of considerable importance. It may be done by calculating the standard error of measurement—written $\sigma_{(M)}$—which may be interpreted as a measure of the amount of variable error, or as a measure of the probable divergence of obtained scores from true scores after the elimination of constant errors. The $\sigma_{(M)}$ is derived directly from the $\sigma_{(\text{est.})}$ as follows. In the equation $\sigma_{(\text{est. }1)} = \sigma_1\sqrt{1-r^2_{12}}$ (see formula 32), if σ_1 is the σ of the scores in test 1, and r_{12} is the correlation between tests 1 and 2, then $\sigma_{(\text{est. }1)}$ measures the accuracy with which individual scores in test 1 may be estimated from a knowledge of the corresponding scores in test 2. Now if the scores on test 2 are taken to represent *true* scores, and the scores on test 1, *obtained* scores on the same test the equation may be written

$$\sigma_{(\text{est. obt.})} = \sigma_{\text{obt.}}\sqrt{1-r^2_{\text{obt. true}}}$$

But $r_{\text{obt. true}} = \sqrt{r_{12}}$, and $r^2_{\text{obt. true}} = r_{12}$ the reliability coefficient.

Hence, substituting these values in the above equation, we have

$$\sigma_{(\text{est. obt.})} = \sigma_1 \sqrt{1 - r_{12}},$$

or writing $\sigma_{(M)}$ for $\sigma_{(\text{est. obt.})}$ finally,

$$\sigma_{(M)} = \sigma_1 \sqrt{1 - r_{12}}. \qquad (63)$$

Formula (63) gives the standard error of measurement for a set of obtained scores. Given r_{12}, the reliability coefficient of the test, and σ_1 (the σ of the test scores) we can, from formula (63) measure the probable divergence of an obtained score from its corresponding true score.

Instead of $\sigma_{(M)}$ we may find $PE_{(M)}$—which is probably more often used—by the formula

$$PE_{(M)} = .6745\sigma_1 \sqrt{1 - r_{12}}. \qquad (64)$$

To illustrate the use of these formulas, suppose that in a group of 100 college men, we obtain an average Army Alpha score of 150 with a σ of 15.00 points; and that the self-correlation of Alpha (found by correlating two forms) is .90. What are the $\sigma_{(M)}$ and $PE_{(M)}$? Applying formula (63), we have

$$\sigma_{(M)} = 15\sqrt{1 - .90} = 4.74$$

and from (64),

$$PE_{(M)} = .6745 \times 15\sqrt{1 - .90} = 3.20.$$

From the $PE_{(M)}$, we may interpret this result to mean that the chances are even that the *true score* of any individual in the group of 100 falls within the range, obtained score$\pm$3.20. For a given obtained score of 175, the chances are even that the true score of this particular man lies within the limits 178.20 and 171.80. Expressed in another way, we may say that 50% of the obtained scores are in error (as compared with their true scores) by not more than $\pm$3.20 points.

In the formulas for $\sigma_{(M)}$ and $PE_{(M)}$, the σ's of the test and its duplicate are assumed to be equal. If this is not at

least approximately true we must write these formulas as follows:

$$\sigma_{(M)} = \frac{(\sigma_1 + \sigma_2)}{2}\sqrt{1 - r_{12}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (65)$$

and

$$PE_{(M)} = .6745 \times \left[\frac{(\sigma_1 + \sigma_2)}{2}\sqrt{1 - r_{12}}\right]. \quad . \quad . \quad . \quad (66)$$

In the illustration above, if the σ obtained from the first form of Alpha, and the σ obtained from the second form of Alpha—had been 15 and 20, respectively, $\sigma_{(M)}$ and $PE_{(M)}$ would be written

$$\sigma_{(M)} = \frac{15+20}{2}\sqrt{1 - .90} = 5.53$$

and

$$PE_{(M)} = .6745 \times 5.53 = 3.73.$$

The student must be careful not to confuse the formulas for $\sigma_{(\text{est.})}$ and $PE_{(\text{est.})}$ with those for $\sigma_{(M)}$ and $PE_{(M)}$. The "estimate" formulas enable us to say with what degree of accuracy we can predict an individual's score on one test,—knowing his score on a second (and usually a different) test. The actual prediction of the "most probable score" is made of course, by means of the regression equation connecting the two tests. The $\sigma_{(M)}$ and $PE_{(M)}$ formulas, on the other hand, enable us to determine the probable divergence of an individual's obtained score from his corresponding true score, when we know (1) the σ and (2) the reliability coefficient of the test.

When tests are scored in different units, the $\sigma_{(M)}$ of the one cannot be directly compared with the $\sigma_{(M)}$ of the other. We cannot compare directly, for example, the reliability of a score made on a tapping test (score in number of taps made in 30 sec.) with the reliability of a score on a logical memory test (scored in number of items remembered). A simple method of overcoming this difficulty is to use a ratio similar to the coefficient of variation, V, described in Chapter I. Thus the ratio

$\frac{\sigma_{(M)}}{\text{Av.}}$ or $\frac{PE_{(M)}}{\text{Av.}}$ of the one test may be compared directly with the $\frac{\sigma_{(M)}}{\text{Av.}}$ or $\frac{PE_{(M)}}{\text{Av.}}$ of the other. In this way, the reliability of obtained scores on one test may be compared with the reliability of the obtained scores on another.

III. Combining the Scores from Different Tests

When a number of different tests have been given to the same individual, it is often desirable be able to combine the separate test scores into a composite score in order to express the individual's standing in the tests as a whole. The simplest procedure is, of course, to average the scores as they stand. In merely averaging results, however, two difficulties arise. The first is the difference in the *size* and *kind* of units employed in the tests. Many tests are given by the Amount-Limit Method—the work is completed (or as much as possible done) and the individual's performance is scored in terms of the time required. Many other tests are given by the Time-Limit Method—the time is fixed, and the subject's score is the number of items completed or the number of questions answered in the time allowed. It is obvious that scores obtained from tests given by these two methods cannot be combined directly.

A second difficulty is the question of the relative influence or "weight" to be given the different tests in the composite score. Simply to average the "raw" (obtained) scores gives us no control over the relative importance of the various tests in the final total score. For although it is often assumed that by simply averaging results we avoid the troublesome question of weighting, what we actually do in such cases is to weight quite drastically without knowing what the weights are. With these two difficulties in mind, let us examine several methods which have been proposed for combining separate test scores into a composite score.

1. Combining Test Scores by Percentiles

If the distribution of each of the separate tests which we have given is broken up into percentiles, it becomes an easy matter to combine the separate percentile rankings in the various tests, and thus secure a final percentile ranking for each individual. The method of calculating percentiles has already been considered (page 45). It is only necessary, then, to show how percentile rankings may be combined.

TABLE XXIX

Percentile Distributions for 9-Year Olds on Three Tests. Method of Combining the Percentile Ratings of a Single Individual

Tests	Percentiles 0	10	20	30	40	50	60	70	80	90	100	*S*'s Score	*S*'s Perc. Rank
Picture Completion	62	240	297	325	372	407	440	450	499	577	646	445	65
Substitution	219	190	173	158	152	141	133	126	121	109	80	126	70
Sequin Form-Board	34	24	21	20	18	18	17	16	15	15	13	17	60
Median percentile													65

Table XXIX gives the percentile tables for 9 year-olds on three tests of the Pintner-Patterson series of performance tests. The subject, a 9 year-old boy, made a score of 445 on Picture Completion which gave him a percentile ranking of 65 (midway between 60 and 70) on this test. On Substitution, a score of 126 gave him a percentile ranking of 70; and on the Sequin Form Board a score of 17 gave him a percentile ranking of 60. The median of these three percentile rankings is 65, which indicates that the subject is somewhat above the average for his age. If the subject had been, say, 10 or 11 years old, percentile tables for these age distributions would have been used. As is evident from Table XXIX the method of combining percentile rankings is simple and straightforward; it rules out the question of different units in the tests combined, and gives each test equal weight in the final score.

2. Combining Test Scores by the Method of Median Mental Age

When the subjects are children, and age-norms exist for the tests administered, it is a relatively easy matter to determine the MA of the subject in each test, and then find the median of these MA's. The median MA is the "composite score."

Tables giving the MA equivalents in scores for various tests have been published by many authors [1] and need not be reproduced here. The method of finding a median mental age for several tests is often very useful and its results are easily interpreted. The method does not, however, apply to normal adults.

3. Combining Tests Which Have Been Weighted According to the Variability of the Test Scores

When several tests have been given, all by the Time-Limit or all by the Amount-Limit Method, scores may be combined directly, the weight which each test score shall have in the composite score being determined in accordance with the variability of the test scores. An illustration will make the method clear. Suppose that in a given test in which the Average$=25$ and $\sigma=5$, subject A scores 20; and in another test in which the Average$=150$ and $\sigma=15$, A scores 160. Now if we simply add A's two scores, e.g., $20+160$ to get 180, the score in the second test is given three times as much importance in this composite as the score in the first, since the spread, i.e., the σ, is three times greater in the second test. In order to give the two tests equal weight, we must equalize their spread or variability, and this can be done by multiplying the σ of the first test by 3 or dividing the σ of the second by 3. This same procedure must then be applied to the scores. By the first operation, our composite score becomes $20\times3+160$ or 220; by the second operation, the

[1] For example, see Whipple, *Manual of Mental and Physical Tests*, Vols. I and II, 1914; Pintner and Patterson, *A Scale of Performance Tests*, 1921; Pyle, W. H., *The Examination of School Children*, 1913.

composite score becomes $20+\frac{160}{3}$ or 73.34. In either composite both tests will now have equal weight.

TABLE XXX

How to Combine Scores Weighed According to Variability

Data from 200 College Women. (From Carothers, F. E., *Psychological Examination of College Students*, Archives of Psychology, 1921, pp. 30–34.)

Tests	Log. Memory (recall) 1	Log. Memory (recognition) 2	Completion 3	Information 4	Vocabulary 5	
Average	6.50	37.47	35.78	104.71	73.90	
σ	1.76	7.69	4.36	26.79	7.60	
Multiplier to give all tests equal weight	5	1	2	$\frac{1}{3}$	1	
New σ	8.80	7.69	8.72	8.93	7.60	
A's score	5	35	30	100	75	
A's weighted score (all tests equal)	25	35	60	34	75	*Total* =229
A's weighted score: Tests 1 and 3 weighted 2, others 1	50	35	120	34	75	=314

In order to illustrate this method of combining scores in more detail, the average and the σ for each of five tests are given in Table XXX together with the scores of subject A on each test. If A's scores are added as they stand, test 4 (Information) will be given 15 times the weight of test 1 (Logical Memory, recall) in the composite, since the σ for Information is 15 times the σ for Logical Memory, recall. Likewise, Information will have approximately 6 times the weight of Completion and approximately 3 times the weight of Logical Memory, recognition, and Vocabulary. It seems hardly probable that Information is as much superior in value as this to the other tests—in fact, it is possibly one of the least important—and hence a new weighting is clearly necessary. The simplest plan at the start will be to weight all of the tests equally as shown in the table. If we multiply the σ of test 1 by 5, the σ of test 2 by 1, the σ of test 3 by 2, the σ of test 4 by $\frac{1}{3}$, and the σ of test 5 by 1, we make all of the σ's approximately equal. Now if we multiply A's scores by

these same "multipliers," the new test scores will all have the same weight in the final composite. In determining multipliers, the best plan is to keep them whole numbers, if practicable, and as small as possible. In Table XXX, for example, the σ's of tests 2 and 5 have been taken as standards because this gives the simplest multipliers for the other tests.

Suppose now that we had wished to give Logical Memory, recall, and Completion *twice* as much weight as the other tests in the composite. To accomplish this we should simply have multiplied the σ's of tests 1 and 3 by 10 and 4 instead of 5 and 2, i.e., we should have multiplied by enough to make their new σ's twice as large as the σ's of the other tests. Of course, when all of the tests have already been weighted 1, we need only double the scores on tests 1 and 3.

To summarize the steps in the method:

(*a*) Find the average and the σ or Q of each test.

(*b*) If the tests are to have equal weight, multiply the σ or Q of each test by factors selected so as to make all of the new σ's or Q's equal. If some tests are to count more heavily than others, make their σ's or Q's proportionally larger.

(*c*) Multiply each S's score by the "multiplier" decided upon in (*b*), and add these new scores. Leave the result as a composite total, or average the new scores if there is some reason for working with smaller numbers.

4. Combining Test Scores by Converting the Scores of Different Tests into Comparable Series

As mentioned above, the chief difficulties in combining the scores of different tests arise from differences in the units in which the tests are scored as well as differences in variability among the tests themselves. We have already considered three ways of avoiding these difficulties. Still another method is to convert the scores of the different tests into comparable distributions, after which the test scores may be combined directly.

Two methods of combining tests in this way have been

proposed, both of which assume that the distributions of test scores are normal or approximately normal. The more recent, suggested by Professor Clark Hull,[1] is to convert the scores from each test into a "standard" normal distribution in which the scores shall range from 0 to 100 with a mean at 50 and σ of 14. [Individual scores rarely spread more than $\pm 3.5\sigma$ above or below the average; hence, since $\frac{50}{3.5} = 14.00$ the σ of this distribution may be taken as 14.00.] Conversion of the scores of a given test is readily made by the following scheme:

Let M = average of the given test.
Let $\sigma = \sigma$ of the given test.
Let X_1 = individual's score on the given test.
Let 50 = average of the converted series.
Let $14 = \sigma$ of the converted series.
Let X = individual's score in the converted series.

Now if $S = \frac{14}{\sigma}$. and $K = 50 - MS$; then $X = K + SX_1$.

To illustrate, suppose that in a given test the average is 16.00, the σ is 3.5, and that subject A scores 18 on the test. What is A's converted score?

$$S = \frac{14}{3.5} \text{ or } 4.00, \text{ and } K = 50 - 16 \times 4 \text{ or } -14.00.$$

Substituting in $X = K + SX_1$, $X = -14 + 4 \times 18 = 58$.

A's score, therefore, in a distribution of Average $= 50$ and $\sigma = 14$ is 58. In other words (assuming a normal distribution), 58 is as far above the average of the distribution whose average is 50, as 18 is above the average of the distribution whose average is 16.00.

An illustration will serve to demonstrate how scores may be combined by this method (Table XXXI).

[1] *The Conversion of Test Scores into Series which shall have any Assigned Mean and Degree of Dispersion*, Journal Applied Psychology, 1922, 6, p. 299.

TABLE XXXI

	Test 1 Word Building	Test 2 Digit Span	Total
Average	16.30	7.4	
σ	4.90	1.3	
A's score	18.00	8.0	
A's converted score	54.86	56.48	55.67

Taking test 1, Word-Building, first, from the formula above, $S=\frac{14}{4.9}$ or 2.86; and $K=50-16.30\times2.86$ or 3.38. Hence, $X=3.38+2.86\ X_1$, and substituting A's score of 18 for X_1 we have $X=54.86$. In like manner, in test 2, Digit Span, $S=\frac{14}{1.3}$ or 10.8; and $K=50-7.4\times10.8$ or -29.92. Accordingly, $X=-29.92+10.8\times8$ (substituting A's score in Digit Span) or 56.48. Averaging A's scores in Word-Building and Digit Span, we have 55.67 as the composite score, which means that A is slightly above average (50) in the two tests.

Since we have computed both K and S for each of the tests, all of the scores on Word-Building may be quickly converted into "new" scores by means of the formula $X=3.38+2.86X_1$; and all of the scores on Digit Span converted into "new" scores by means of the formula $X=-29.92+10.8X_1$. In each case the X_1 represents the actual score on the test.

An earlier method of combining test scores, based on the same principles as the above plan, was outlined in 1912 by Professor Woodworth.[1] Woodworth's plan was to find the difference between a given individual's score on a test and the average score, i.e., $X-\mathrm{Av}_x$; divide this plus or minus difference ($\pm x$) by the σ of the test and call the result $\left(\frac{\pm x}{\sigma_x}\right)$, the "reduced score."[2] Reduced scores found in this way for the

[1] *Combining the Results of Several Tests, A Study in Statistical Method,* Psychological Review, 1912, Vol. XIX, pp. 97–123.

[2] Note that in Woodworth's method the average is taken at 0 and σ as 1.00.

same individual on several tests may be combined by simply averaging them—the weight of each test in the composite will be 1.00. To illustrate the method using the data of Table XXXI, A's score of 18 on the Word-Building test is 1.70 above the average, i.e., above 16.30; and dividing this deviation by the σ of the series gives A a "reduced score"—a score expressed in σ units—of .347. On the Digit Span test, A's score of 8.00 is .6 above the average of the distribution, i.e., above 7.4; and dividing .6 by 1.3 we get a reduced score on Memory Span of .462. If we average these two reduced scores, A is found to stand .405 (in σ units) *above* the average of the group in the two tests. (Remember that this method, like the preceding one, assumes that the distributions of test scores are approximately normal.)

Of these two methods, the first is somewhat the simpler inasmuch as it involves only plus values (all transmuted scores lie between 0 and 100), while the second method introduces plus and minus values which are nearly always fractions, often small in size and inconvenient to handle. Again, a composite score of 55.67 by Hull's method is probably more intelligible to the average student accustomed to think in per cents, than an average score of .405 found by Woodworth's plan. The latter result is meaningful only to those who have had considerable statistical training.

Woodworth's method has one particular advantage, however, which should be mentioned, viz., that when reduced scores have once been calculated for two or more tests, correlations between the tests may easily be found. The method of obtaining such correlations is illustrated in Table XXXII which gives the reduced scores made by 10 adults on a Memory Span and Information test, and the correlation between the two series. As shown in the table the calculations are relatively simple. Since each individual's reduced score on Memory Span (X) is simply his x (i.e., his deviation from the average) divided by σ_x, and his reduced score on Information (Y) is, again, his y (i.e., deviation from the average) divided by σ_y, the sum of the

products $\left(\text{i.e., } \frac{x}{\sigma_x} \cdot \frac{y}{\sigma_y}\right)$ of the reduced scores of all of the ten individuals will give $\frac{\Sigma xy}{\sigma_x \sigma_y}$. We know from formula (24) that $r = \frac{\Sigma xy}{N\sigma_x\sigma_y}$ (page 168). Hence, the correlation between the two tests is obtained simply by dividing $\frac{\Sigma xy}{\sigma_x\sigma_y}$, (7.31) by N (10): that is, r equals .731.

TABLE XXXII

To Illustrate the Method of Finding Correlation from "Reduced Scores"

Individuals	Memory Span (X) Score	Information (Y) Score	Reduced Score in X $\left(\frac{x}{\sigma_x}\right)$	Reduced Score in Y $\left(\frac{y}{\sigma_y}\right)$	Product of Reduced Scores $\left(\frac{xy}{\sigma_x\sigma_y}\right)$
A	5	90	−1.19	......	......
B	9	60	.39	−1.45	−.566
C	8	90	......	......	......
D	7	85	− .39	− .24	.094
E	6	70	− .79	− .97	.766
F	10	100	.79	.49	.387
G	12	130	1.58	1.94	3.065
H	6	80	− .79	− .49	.387
I	5	75	−1.19	− .73	.869
J	12	120	1.58	1.46	2.307

$$\frac{\Sigma xy}{\sigma_x\sigma_y} = 7.309$$

$Av_x = 8.0$ $\qquad$ $Av_y = 90.00$

$\sigma_x = 2.53$ $\qquad$ $\sigma_y = 20.62$

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y} = \frac{7.31}{10} = .731$$

Note.—This table is intended simply to illustrate the method. A product-moment r would not ordinarily be found for 10 cases.

The student should bear in mind when using either of these methods that neither is strictly applicable when the distributions are considerably skewed. As stated above, both assume that the distributions to which they are applied are normal or approximately normal.

IV. The σ of the Sum or Difference of Corresponding Values of Two Series of Test Scores

If we know the correlation between two series of test scores X_1 and X_2 and the σ's of the two series, it is possible to compute, in a simple way, the σ of the new composite series obtained by *adding* or *subtracting* the corresponding scores in the two original series. When the scores of the "new" distribution have been found by *adding* corresponding scores, the formula for σ_s[1] is

$$\sigma_s = \sqrt{\sigma^2_{x_1} + \sigma^2_{x_2} + 2r\sigma_{x_1}\sigma_{x_2}}, \quad . \quad . \quad . \quad . \quad . \quad (67)$$

in which σ_s denotes the σ of the "new" *summed*-series, σ_{x_1} is the σ of the X_1 scores, σ_{x_2} is the σ of the X_2 scores, and r is the coefficient of correlation between X_1 and X_2. When the scores in the new distribution have been obtained by *subtracting* corresponding scores in the two tests, formula (67) becomes

$$\sigma_d = \sqrt{\sigma^2_{x_1} + \sigma^2_{x_2} - 2r\sigma_{x_1}\sigma_{x_2}}, \quad . \quad . \quad . \quad . \quad . \quad (68)$$

in which σ_d is the σ of the new *difference*-series.

A problem will illustrate the use of these formulas. Let X_1 denote a Verb-Object Test and X_2 an Opposites Test. Then given $\sigma_{x_1} = 11.18$, $\sigma_{x_2} = 9.00$, and $r_{x_1x_2} = .60$, what is the σ of the new series obtained (1) by *adding* the corresponding X_1 and X_2 scores, and (2) by *subtracting* the corresponding X_1 and X_2 scores? Substituting in formula (67), we have

$$\sigma_s = \sqrt{(11.18)^2 + (9.00)^2 + 2 \times .60 \times 11.18 \times 9},$$

or

$$\sigma_s = 18.07.$$

Thus, 18.07 is the σ of the $(X_1 + X_2)$ series. To find the σ of the $(X_1 - X_2)$ series, σ_d, we substitute in formula (68),

$$\sigma_d = \sqrt{(11.18)^2 + (9.00)^2 - 2 \times .60 \times 11.18 \times 9.00},$$

or

$$\sigma_d = 9.23.$$

[1] For a simple mathematical proof of this formula, see Yule, *An Introduction to the Theory of Statistics*, pp. 210–211.

Formula (68) is often useful when a test has been repeated in a group under changed conditions and the variability of these changes, i.e., the σ of the differences between scores made on the second and the first giving of the test, is sought. Except that there is only the one test concerned, the method is identical with that of the problem above. The chief objection to the formula is that the r between the scores on the first and second giving of the test must be known. For this reason, unless the r is wanted for other purposes, it is usually easier to subtract the corresponding scores and derive the σ of their differences directly.

From the formula for the reliability of the average, $\sigma_{av.} = \frac{\sigma_{(dis.)}}{\sqrt{N}}$, (formula 13), we know that $\sigma_{(dis.)} = \sqrt{N}\sigma_{av.}$. We may, therefore, write $\sqrt{N}\sigma_{av.\,x_1}$ instead of σ_{x_1}; $\sqrt{N}\sigma_{av.\,x_2}$ instead of σ_{x_2}; $\sqrt{N}\sigma_{av.\,s}$ instead of σ_s; and $\sqrt{N}\sigma_{av.\,d}$ instead of σ_d. Making these substitutions in formulas (67) and (68) we have (the N's cancel), that

$$\sigma_{av.\,s} = \sqrt{\sigma^2_{av.\,x_1} + \sigma^2_{av.\,x_2} + 2r\sigma_{av.\,x_1}\sigma_{av.\,x_2}}, \quad . \quad . \quad (69a)$$

and

$$\sigma_{av.\,d} = \sqrt{\sigma^2_{av.\,x_1} + \sigma^2_{av.\,x_2} - 2r\sigma_{av.\,x_1}\sigma_{av.\,x_2}}, \quad . \quad . \quad (69b)$$

in which $\sigma_{av.\,s}$ is the σ of the average of the (X_1+X_2) series of scores, and $\sigma_{av.\,d}$ is the σ of the average of the (X_1-X_2) series of scores.

Formulas (69a) and (69b) must always be used whenever there is any *correlation* between the X_1 and X_2 scores. If X_1 and X_2 are *uncorrelated*, that is, if $r = .00$, the third term under the radical disappears and (69a) and (69b) become

$$\sigma_{av.\,s} = \sqrt{\sigma^2_{av.\,x_1} + \sigma^2_{av.\,x_2}}, \quad . \quad . \quad . \quad . \quad (70a)$$

and

$$\sigma_{av.\,d} = \sqrt{\sigma^2_{av.\,x_1} + \sigma^2_{av.\,x_2}}. \quad . \quad . \quad . \quad . \quad (70b)$$

Now if we write $\sigma_{(diff.)}$ instead of $\sigma_{av.\,d}$ in formula (70b), we at once recognize the familiar formula, $\sigma_{(diff.)} = \sqrt{\sigma^2_{av.\,1} + \sigma^2_{av.\,2}}$,

which we have used heretofore for measuring the reliability of the difference between two averages, or with appropriate changes, two σ's, or two r's. It should always be remembered that $\sigma_{(\text{diff.})}$ is simply a special form of the more general formula (69*b*) and that it always assumes a *zero correlation* between X_1 and X_2.

The *PE* may be written for σ in any of the formulas given in this Section by making the substitution $PE = .6745 \times \sigma$.

V. How to Interpret the Coefficient of Correlation between Two Tests

When can a coefficient of correlation be considered "high"? Is an r of .40 between two tests evidence of "low" or "marked" relationship? Questions like these, and many others which relate to the interpretation of a coefficient of correlation frequently arise in test work and must be answered if we would understand the significance of an obtained r.

The effectiveness of an r as a measure of relation may be evaluated in several ways: (1) in terms of the standard error of estimate; (2) in terms of the standard error of measurement; and (3) in terms of the percentage of factors common to the two capacities correlated. Let us consider these three approaches to an interpretation of r before attempting to lay down any general rule for classifying r's as "high," "medium," or "low."

1. The Interpretation of a Coefficient of Correlation in Terms of $\sigma_{(\text{est.})}$

The standard error of estimate, $\sigma_{(\text{est.})}$, is probably the most practicable way of evaluating the effectiveness of a coefficient of correlation. This follows from the fact that $\sigma_{(\text{est. } X_1)}$, which enables us to tell how accurately we can estimate an individual's score on test X_1 knowing his score on test X_2, depends on the r between the two tests. When $r = 1.00$, $\sigma_{(\text{est. } X_1)} = .00$, which means that we can predict a score in X_1 from a knowledge of X_2 with perfect accuracy—no error.

To take the opposite extreme, when $r = .00$, $\sigma_{(\text{est. } X_1)} = \sigma_1$ directly, which means that we can only be certain that the predicted score lies somewhere within the limits of the X_1 distribution, i.e., within the limits, Obtained Score $\pm 3\sigma$. In other words, the estimate from the distribution of X_1 alone is as good as the estimate made with the addition of X_2. As r decreases from 1.00 to 0, the standard error of estimate rapidly increases, so that predictions from the regression equation range all of the way from certainty to practically guesswork. The closeness of the correspondence denoted by an r, therefore, may be gauged by the size of $\sigma_{(\text{est.})}$.

We may illustrate with the following problem. Suppose that the correlation between two tests X_1 and X_2 is .60, and that $\sigma_{x_1} = 5.00$. Then $\sigma_{(\text{est. } X_1)}$ is $5 \times \sqrt{1 - .6^2}$ or 4.00, which is only 20% less than 5.00 the $\sigma_{(\text{est. } X_1)}$ for $r = .00$, i.e., for a minimum predictive value. The proportionate amount of reduction in $\sigma_{(\text{est. } X)_1}$ as r varies from .00 to 1.00 is given by the expression $\sqrt{1 - r^2}$, and hence it is possible to estimate the "predictive" value of an r from $\sqrt{1 - r^2}$ alone. This radical ($\sqrt{1 - r^2}$) has been designated by Kelley[1] the "coefficient of alienation," and is usually denoted by the letter "k." k may be thought of as measuring the *absence* of relation between two variables X_1 and X_2, in the same way that r measured the *presence* of relation. Thus when $k = 1.00$, $r = .00$, and when $k = .00$, $r = 1.00$—the larger the coefficient of alienation the greater the lack of relation, and the less the value of the prediction. In order to show how the estimate improves as r increases, the k's for the values of r from .00 to 1.00 are given in Table XXXIII.

It will be noted that r must be .866 before k is half way between perfect correlation, and a guess:—before the standard error of estimate is reduced one-half. For r's of .30 and less, the coefficients of alienation are so large that the predic-

[1] Kelley, T. L., *Principles Underlying the Classification of Men.* Journal of Applied Psychology, 1919, Vol. III, 1, p. 50.

tions based on them are but little better than a guess. Even with an $r = .99$, it will be noticed that the standard error of estimate is still $\frac{1}{7}$ as large as when $k = 1.00$. It is obvious, then, that in order to estimate individual scores with accuracy, the correlation should be at least .90.

TABLE XXXIII

GIVING COEFFICIENTS OF ALIENATION k FOR VALUES OF r FROM .00 TO 1.00

r	$k = \sqrt{1 - r^2}$	r	$k = \sqrt{1 - r^2}$
.00	1.0000	.80	.6000
.10	.9950	.8660	.5000
.20	.9798	.90	.4539
.30	.9539	.95	.3122
.40	.9165	.98	.1990
.50	.8660	.99	.1411
.60	.8000	1.00	.0000
.70	.7141		
(.7071)	.7071		

2. The Interpretation of a Coefficient of Correlation in Terms of the Standard Error of Measurement, $\sigma_{(M)}$.

We have found (page 183) that the standard error of measurement enables us to estimate the probable divergence of an obtained score on a test from its corresponding true score. Moreover, since $\sigma_{(M)} = \sigma_1 \sqrt{1 - r_{12}}$, the amount of this probable divergence will depend to a large degree upon the size of the self-correlation, r_{12}, and accordingly it follows that the value of r_{12} as a measure of relation may be determined from the size of $\sigma_{(M)}$. When $r = 1.00$, for example, $\sigma_{(M)} = .00$, and every obtained score equals its true score exactly. When $r = .00$, on the other hand, $\sigma_{(M)} = \sigma_1$ (the σ of the distribution) and we can only be sure that the true score (corresponding to a given obtained score) lies somewhere within the limits of the distribution—within the limits $\pm 3\sigma$. In other words, when $r = .00$, the probable divergence of an obtained score from its true score is as great as it would be had we simply guessed that the true score lay somewhere in the distribution.

To illustrate, suppose that the reliability coefficient of a given

test, $r_{12} = .80$, and that $\sigma_1 = 10.00$. Then $\sigma_{(M)} = 10\sqrt{1-.80}$ or 4.472, and since $\sigma_{(M)}$ is 10.00 when $r = .00$, evidently a reliability coefficient of .80 serves to reduce $\sigma_{(M)}$ to about 45% of what it would be in the event of a guess. The reduction in $\sigma_{(M)}$ as r varies from 0 to 1.00 is given by the expression $\sqrt{1-r_{12}}$. Hence this factor may be used to test the effectiveness of an obtained reliability coefficient, just as k tests the value of the r between two tests. In Table XXXIV the values of $\sqrt{1-r_{12}}$ have been calculated for r's from .00 to 1.00.

TABLE XXXIV

GIVING VALUES OF $\sqrt{1-r_{12}}$ FOR VALUES OF r FROM .00 TO 1.00

r	$\sqrt{1-r_{12}}$	r	$\sqrt{1-r_{12}}$
.00	1.0000	.80	.4472
.10	.9487	.90	.3162
.20	.8944	.95	.2236
.30	.8367	.98	.1414
.40	.7746	.99	.1000
.50	.7071	1.00	.0000
.60	.6325		
.70	.5477		
.75	.5000		

From Table XXXIV it is evident that the self-correlation of a test must be at least .75 before $\sqrt{1-r_{12}}$ is half way between complete reliability and a guess. For an $r_{12} = .98$, the chances are still 68 in 100 that a given score will diverge from its true score by as much as $\pm .1414$ of the σ of the test. Since high reliability coefficients, therefore (e.g., .90 or above), indicate relatively large departures from perfect reliability, it is clear that a self-correlation of, say, .30 or .40 is almost valueless.

3. Interpretation of a Coefficient of Correlation in Terms of the Percentage of Common (Overlapping) Elements or Factors

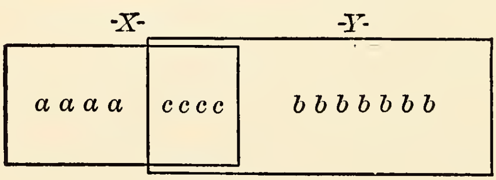
It is sometimes helpful to regard a coefficient of correlation as a ratio which expresses—directly or indirectly—the per-

centage of elements or factors common to the tests which are correlated. Or again, r may be thought of as a device for indicating the extent to which the factors which determine capacity in the one test "overlap" those of another test.[1] Let us suppose that capacity in test X depends upon the presence or absence of $a+c$ independent, elemental, factors; and that capacity in test Y depends upon the presence or absence of $b+c$ independent, elemental, factors. The a factors determine X scores alone, the b factors Y scores alone, and the c factors are common to both X and Y. Moreover, let us suppose further that all factors, a, b, and c, are governed solely by the laws of chance, so that each factor is as likely to be present as absent in the same way that a coin when tossed is as likely to fall heads as tails.

Now if we let n_a = total number of a factors, n_b = total number of b factors, and n_c = the total number of c factors, it can be shown [2] that the correlation between X and Y is given by the formula:

$$r = \frac{n_c}{\sqrt{(n_a + n_c)(n_b + n_c)}} \quad . \; . \; . \; . \; . \; . \; . \quad (71)$$

That is, the coefficient of correlation equals the number of common factors in X and Y, divided by the geometrical mean of the total number of factors in X and Y. This situation is shown graphically in Diagram XXVII in which X is determined by 8 factors, 4 a's and 4 c's, and Y by 11 factors, 7 b's and 4 c's. The correlation by formula (71) is



$$r = \frac{4}{\sqrt{8 \times 11}} = .426$$

DIAGRAM XXVII

$$\frac{4}{\sqrt{(4+4)(7+4)}} \text{ or } \frac{4}{\sqrt{8 \times 11}} = .426$$

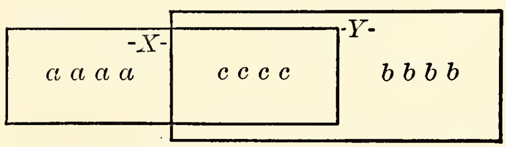
[1] The following is adapted from the discussion by Kelley, *Statistical Method*, pp. 189–190.

[2] See Kelley, *Statistical Method*, 1923, p. 190; or Brown, Wm., *Essentials of Mental Measurement*, 1911, pp. 79–80.

If the number of elementary factors determining the score in X equals exactly the number determining the score in Y, so that $n_a = n_b$, formula (71) becomes

$$r = \frac{n_c}{n_a + n_c}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (72)$$

and the coefficient of correlation is now simply the decimal fraction which indicates what proportion of the causes influencing performance in X and Y are common to both. If t = the number of common factors (n_c) and if s = the total number of factors, present in X and Y ($n_a + n_c$) r is simply $\frac{t}{s}$. (Remember that the factors in X and Y are assumed to be equal in number and influence.) This condition is illustrated in Diagram XXVIII. Since X is determined by 8 factors, 4 a's and 4 c's and Y by 8 factors, 4 b's and 4 c's, the correlation by formula (72) is 4/8 or .50.



$$r = \frac{4}{8} = .50$$

DIAGRAM XXVIII

Now let us assume, lastly, that Y is *completely* determined by n_c elements, and that X is determined by these same elements *plus* n_a elements in addition ($n_b = 0$). Formula (71) then becomes

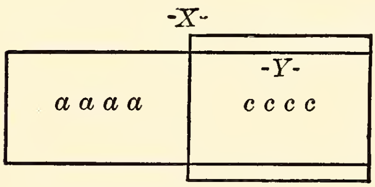
$$r = \frac{n_c}{\sqrt{n_c(n_a + n_c)}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (73)$$

and the coefficient of correlation equals the number of common elements in X and Y divided by the geometrical mean of the total number of factors in X and in Y. Diagram XXIX shows this graphically. Y is determined by 4 c's and X by these factors plus 4 a's in addition: the correlation, therefore, is $\frac{4}{\sqrt{4 \times 8}}$ or .707. If

we square the r obtained from formula (73), we have that

$$r^2 = \frac{n_c}{n_a + n_c}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (74)$$

that is, the square of the coefficient gives the extent to which the elements in Y overlap those of X:—or the proportion of elements in X which are also involved in Y. In Diagram XXIX note that Y overlaps X 50% and that r^2—i.e., $(.707)^2$—is .50 as it should be.[1] Moreover, since the coefficient of alienation will equal .707 when $r = .707$ (see Table XXXIII), it follows that an r of .707 (and not .50) should be taken as *half* of a perfect correlation.[2] On the same assumptions, an overlapping of $33\frac{1}{3}\%$ common elements—i.e., $r^2 = .3334$—will give a correlation of .578, which is 1/3 of a perfect correlation; and an overlapping of 25% common elements, $r^2 = .25$, gives an $r = .50$, which is 1/4 of a perfect correlation. By analogy, an r of .30 or less implies so slight a degree of overlapping that there can be a very small percentage of common elements.



$$r = \frac{4}{\sqrt{4 \times 8}} = .707$$

DIAGRAM XXIX

The coefficient of correlation as a measure of the percentage of common factors may be seen to best advantage in series formed by tossing coins or throwing dice, in which the "overlapping" is arbitrarily determined and controlled at will. As an illustration, consider the correlation table in Diagram XXX in which is shown the relation between two series of 500 successive throws of 12 pennies made in the fol-

[1] This result has interesting implications. Thus if all of the elements in test X_2 are common to X_1 (e.g., a criterion) the extent to which X_2 overlaps X_1 is given by simply squaring the coefficient, $r_{x_1x_2}$. The assumption must be made, of course, that the scores in both tests are summations of independent and similar elements whose presence or absence is governed by chance alone.

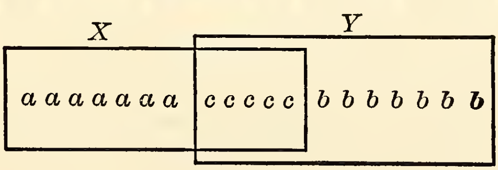
[2] Woodworth, R. S., *Combining the Results of Several Tests: A Study in Statistical Method*, Psychological Review, 1912, XIX, p. 113. Hull Clark, *The Joint Yield from Teams of Tests*, Journal of Educational Psychology, 1923, 14, pp. 396–406.

DIAGRAM XXX

Showing the number of heads in 500 successive throws of 12 pennies in which 7 pennies were tossed in the second throw and 5 remained as they fell in the first throw of all 12 together.

HEADS IN FIRST TOSS

HEADS IN SECOND TOSS	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
12														
11								1						1
10					1		2		2	3	1	1		10
9						2	9	13	4	3				31
8				1	5	9	10	18	14	4	2			63
7			1	2	5	14	24	28	10	7	4			95
6			1	3	9	18	27	29	16	3	2	1		109
5				4	11	23	21	15	9					83
4			3	6	9	21	14	10	5	1				69
3			3	3	8	4	4	4						26
2			3	1	5	1	1							11
1					1	1								2
0														
Total			11	20	54	93	112	118	60	21	9	2		500



$n_c = 5$

$n_a = n_b = 7$

$$r = \frac{n_c}{n_a + n_c} = \frac{5}{12} = .416 \qquad (72)$$

By calculation (product-moment)

$r = .424.$

[1] From Pearl, R., *Medical Biometry and Statistics*, p. 297 (after Darbishire).

DIAGRAM XXXI

Showing the results of 100 successive throws of dice in first throw of which (X) 5 dice were thrown, counted, and left down; and in each second throw of which (Y) 5 additional dice were thrown and counted together with the 5 left down (10 in all).

First Throw of 5 Dice (X)

Second Throw of 10 Dice (Y)	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total
45													1	1	1	3
44									1							1
43								2				1		2		5
42								1	1	1		1	1		1	6
41										1			1			2
40									2	3			1			6
39									2	2	1			1		6
38							1		1	1	2	1				6
37			1					1	2	2	1	1			1	9
36					1	1	2	2	1	1						8
35					1		1	2	1	1	2		1			9
34			1	1			2	1	1	2		1				9
33	1			1					1							3
32						1	2	1				1				5
31			1	1		1			1	1						5
30					2		1	1								4
29	1							1								2
28			2		1	1	1									5
27			2				1									3
26	1															1
25		1			1											2
Total	3	1	7	3	6	4	11	12	14	15	6	6	5	4	3	100

By calculation (product-moment)

$r = .694$

$n_c = 5 \qquad n_a = 5$

—X—

(5)	(5)
$a\ a\ a\ a\ a$	$c\ c\ c\ c\ c$

—Y—

$$r = \frac{n_c}{\sqrt{n_c(n_a + n_c)}} = \frac{5}{\sqrt{5 \times 10}} = .707 \qquad (73)$$

lowing way: first, all 12 pennies were tossed, and the number of heads recorded and noted in the X column; then 5 coins were left lying and the remaining 7 were tossed again and the number of heads in all 12 recorded and noted in column Y, opposite the X entry. By this scheme 5 coins (factors) contribute to each pair of tosses; and hence, according to formula (72) the correlation should be 5/12 or .416. By the product-moment formula the actual correlation between the two series is .424, which indicates a very close correspondence between actual and theoretical results. The situation existing in each pair of X and Y tosses is shown in the figure in Diagram XXX. If 4 coins had been left lying, the r would have been 4/12 or .334; if 6 had been left lying, r would have been 6/12 or .50 etc. A number of diagrams of the sort shown, in which the number of common factors (i.e., coins left lying) varies from 0 to 12, and r from 0 to 1.00 may be found in Pearl's *Medical Biometry and Statistics*, pages 294–300.

Now suppose that we calculate the correlation between two series of dice throws made according to the following scheme:[1] 5 dice are thrown, and the total read and recorded in the X column; then 5 additional dice are thrown and the total of *all* 10 (the 5 left and 5 just thrown) are read and recorded in the Y column. If this is continued until 100 throws have been made, we shall have 100 X and 100 Y entries, each Y throw (of 10 dice) "overlapped" to the extent of 50% by its corresponding X throw (of 5 dice). And since all of the elements in X are completely contained in Y, the correlation between X and Y should, by formula (73), be $\frac{5}{\sqrt{5 \times 10}}$ or .707. (See Diagram XXXI and accompanying figure.) Actually, the correlation by the product-moment formula is .694, which indicates, again, a very close correspondence between actual and theoretical results. The square of this r gives us approximately .50 as the percentage of common elements in X and Y:

[1] These throws were made by the writer.

that is, we have one half of a perfect correlation. (See page 294.)

While formulas (71–74) are interesting and suggestive as giving us the means of interpreting a coefficient of correlation under certain special or restricted conditions, it would be a mistake to apply them generally,—to assume that by simply squaring the coefficient of correlation we can always determine the percentage of common factors or the amount of overlapping. It seems likely that the scores on most psychological tests as well as many social and educational measurements are the result of the combined action of many factors which are often dependent on each other, and probably interwoven in a relatively complex manner. At any rate, we do not know that a test score is simply the sum of a certain number of similar and independent elements.

Summary

From the discussion in the preceding paragraphs, it is evident that even with correlation coefficients which we have been accustomed to think of as high, the departure from perfect correlation is considerable. Strictly speaking, the term "high correlation" should be applied only to coefficients which are .95 or above. However, in mental, social, and educational measurements there are so many actual and potential sources of error due to the variability of the material dealt with, and the relative crudity of the measurements made, that very few tests indeed could meet this requirement. Very seldom do correlations between tests run above .70 or .75; and hence it is probably justifiable, in view of the limitations mentioned, to regard such coefficients as high. There seems to be fairly general agreement among workers with tests that an

r from .00 to $\pm$ .20 denotes indifferent or negligible relation.
r from $\pm$.20 to $\pm$ .40 denotes low correlation: present but slight.
r from $\pm$.40 to $\pm$ .70 denotes substantial or marked relationship.
r from $\pm$.70 to $\pm$1.00 denotes high relation.

This is a tentative classification which is to be taken as only

generally true. The size of a correlation coefficient should always be evaluated with due regard for the material dealt with, the size of the sample, and PE_r, no matter what its absolute value.

PROBLEMS

1. The self-correlation of a certain test is .60.
   (*a*) How much must the test be lengthened to raise the self-correlation to .90?
   (*b*) What effect will doubling the test have on its reliability?

2. Two equivalent half-scales are made up from the Downey Will-Temperament [1] Test in the following way: (1) by grouping all odd-numbered tests in one half-scale, and all even-numbered tests in the other; (2) by grouping the first two tests of every pattern into one half-scale, and the last two tests into another half-scale; (3) by grouping the first and last tests of each pattern into one half-scale, and the second and third tests of each pattern into a second half-scale.

   Reliability coefficients for the half-scale were found as follows by the three methods:

Method	Reliability Coefficient	$N = 146$
1	.17	
2	.31	
3	.24	
Average	.24	

   What is the reliability of the whole Downey test?

3. In a small group the reliability coefficient of a test is .55 and the σ of the test scores is 3.00. What must the self-correlation of this test be in a larger group whose σ is 5.00, in order to have the same degree of reliability?

4. The reliability coefficient of a test, as found in a large unselected group, is .92; the Average is 142 and σ is 16.00. If an individual makes 150 on the test,
   (*a*) What is the PE of this score, i.e., the $PE_{(M)}$?
   (*b*) Within what range does the true score lie?

[1] Ruch, G. M., and Del Manzo, M. C., *The Downey Will-Temperament Group Test: A Further Analysis of Its Reliability and Validity.* Journal Applied Psychology, Vol. VII, 1923, p. 65.

(*c*) In a second test of a different function, the reliability coefficient is .86; the average is 54 and σ is 10.00. In which test are the obtained scores the more reliable, i.e., closer to the true scores?

5. The reliability coefficient of a test is .80. What is the maximum self-correlation obtainable with this test as it stands?

6. Given the following records (all in *seconds*) for 100 Barnard Freshmen;[1] and the scores made by individual *A*.

Tests	Coordinate	Tapping	Color Naming	Opposites
Average............	82.7	376.3	57.0	51.1
SD.................	10.8	51.7	8.8	10.3
A's scores...........	85	350	62	40

(*a*) Combine *A*'s scores by the method of variability weighting all tests 1.

(*b*) Combine *A*'s scores weighting Coord. and Tapping 1 each, Color Naming 3, and Opposites 4.

7. Using the data in Example 6 above, combine *A*'s scores by the two methods given on pages 282 and 283. Since all scores are in seconds, the higher the score numerically the lower it actually is.

8. One hundred and fifty high school seniors make an average score of 120 on Army Alpha with a σ of 21.6. Two weeks later the group is praised for its performance (without, however, being told what the scores were) and given a second form of Alpha on which the average score is 126 and the σ is 24.2. The r between the tests is .86.

(*a*) Is the effect of the incentive (praise) plus the practice effect sufficient to bring about a real increase in average score? How would you rule out the practice effect?

(*b*) Why is it necessary to have the correlation between the tests?

9. A battery of tests correlates .85 with a criterion. Assuming that performance on the battery is completely determined by X elements, and performance on the criterion by $X+Y$ elements, to what extent may we say that the battery probably "overlaps" the criterion?

10. Interpret a coefficient of correlation $r=.50$ in three ways; an $r=.65$?

[1] Carothers, F. E., *The Psychological Examination of College Students*, Archives of Psychology, 1921, No. 46, pp. 21ff.

Answers

1. (*a*) 6 times.
   (*b*) $r = .75$
2. Method 1: $r = .29$. Method 2: $r = .47$. Method 3: $r = .39$. Average of all three methods: $r = .38$.
3. $r = .84$.
4. (*a*) $PE_{(M)} = 3.05$.
   (*b*) Between 162.2 and 137.8.
   (*c*) In the first test. The $\frac{PE_{(M)}}{\text{Av.}} = .021$ (first test); $\frac{PE_{(M)}}{\text{Av.}} = .047$ (second test).
5. $r = .89$.
6. (*a*) Taking as multipliers for the four tests, 1, $\frac{1}{5}$, 1, and 1, respectively, we have 257 as A's composite score.
   (*b*) A's score is 501. (Since the measures of performance are in time units, the higher the numerical score the lower the actual performance.)
7. A's scores are 47, 57, 42, and 65. Her average is 52.75. (Hull's method.)
   A's scores are $-.213$, $+.509$, $-.568$, $+1.078$; her average is .202. (This means that A stands $.202\sigma$ above the average cf the group on the four tests.)
8. (*a*) Yes. $\frac{D}{\sigma_{\text{diff.}}}$ is 5+.
9. About 72% common elements.

REFERENCES

The following books will be found to be helpful as general references:

1. Primer of Statistics, by W. P. and E. M. Elderton. A. & C. Black, Ltd., London. 1910.
2. Mental and Social Measurements, by Edward L. Thorndike. Published by Teachers College, Columbia University. 1912 (revised edition).
3. Statistical Methods Applied to Education, by Harold O. Rugg. Houghton Mifflin Company. 1917.

4. An Introduction to Statistical Methods, by Horace Secrist. Macmillan Company. 1917.
5. How to Measure in Education, by Wm. M. McCall. The Macmillan Company. 1922.
6. The Theory of Educational Measurements, by Walter Scott Monroe. Houghton Mifflin Company. 1923.
7. The Fundamentals of Statistics, by L. L. Thurstone. The Macmillan Company. 1925.
8. Statistical Method in Educational Measurement, by Arthur S. Otis. World Book Company. 1925.

More advanced books are:

1. Elements of Statistics, by A. L. Bowley. P. S. King and Son, London. 1920 (fourth edition).
2. An Introduction to the Theory of Statistics, by G. Udny Yule. Chas. Griffin and Company, London. 1919 (5th edition).[1]
3. Essentials of Mental Measurement, by W. M. Brown and G. H. Thomson. Cambridge University Press. 1920.
4. A First Course in Statistics, by D. Caradog Jones. G. Bell & Sons, London. 1921.
5. Statistical Method, by Truman L. Kelley. The Macmillan Company. 1923.
6. Handbook of Mathematical Statistics, by H. L. Rietz et al. Houghton Mifflin Company. 1924.

Aids to Computation:

1. Barlow's Tables of Squares, Cubes, Square Roots, Cube Roots, Reciprocals of numbers from 1 to 10,000. E. and F. N. Spon, Ltd., London. 1921.
2. Tables of $\sqrt{1-r^2}$ and $1-r^2$ for use in Partial Correlation and Trigonometry, by John Rice Miner, Sc.D. Johns Hopkins Press. 1922.

[1] The book by Yule is a classic which should be known to every serious student of mental and social measurements.

Table of Squares and Square Roots of the Numbers from 1 to 1000

Number	Square	Square Root	Number	Square	Square Root
1	1	1.000	51	26 01	7.141
2	4	1.414	52	27 04	7.211
3	9	1.732	53	28 09	7.280
4	16	2.000	54	29 16	7.348
5	25	2.236	55	30 25	7.416
6	36	2.449	56	31 36	7.483
7	49	2.646	57	32 49	7.550
8	64	2.828	58	33 64	7.616
9	81	3.000	59	34 81	7.681
10	1 00	3.162	60	36 00	7.746
11	1 21	3.317	61	37 21	7.810
12	1 44	3.464	62	38 44	7.874
13	1 69	3.606	63	39 69	7.937
14	1 96	3.742	64	40 96	8.000
15	2 25	3.873	65	42 25	8.062
16	2 56	4.000	66	43 56	8.124
17	2 89	4.123	67	44 89	8.185
18	3 24	4.243	68	46 24	8.246
19	3 61	4.359	69	47 61	8.307
20	4 00	4.472	70	49 00	8.367
21	4 41	4.583	71	50 41	8.426
22	4 84	4.690	72	51 84	8.485
23	5 29	4.796	73	53 29	8.544
24	5 76	4.899	74	54 76	8.602
25	6 25	5.000	75	56 25	8.660
26	6 76	5.099	76	57 76	8.718
27	7 29	5.196	77	59 29	8.775
28	7 84	5.292	78	60 84	8.832
29	8 41	5.385	79	62 41	8.888
30	9 00	5.477	80	64 00	8.944
31	9 61	5.568	81	65 61	9.000
32	10 24	5.657	82	67 24	9.055
33	10 89	5.745	83	68 89	9.110
34	11 56	5.831	84	70 56	9.165
35	12 25	5.916	85	72 25	9.220
36	12 96	6.000	86	73 96	9.274
37	13 69	6.083	87	75 69	9.327
38	14 44	6.164	88	77 44	9.381
39	15 21	6.245	89	79 21	9.434
40	16 00	6.325	90	81 00	9.487
41	16 81	6.403	91	82 81	9.539
42	17 64	6.481	92	84 64	9.592
43	18 49	6.557	93	86 49	9.644
44	19 36	6.633	94	88 36	9.695
45	20 25	6.708	95	90 25	9.747
46	21 16	6.782	96	92 16	9.798
47	22 09	6.856	97	94 09	9.849
48	23 04	6.928	98	96 04	9.899
49	24 01	7.000	99	98 01	9.950
50	25 00	7.071	100	1 00 00	10.000

Table of Squares and Square Roots—*Continued*

Number	Square	Square Root	Number	Square	Square Root
101	1 02 01	10.050	151	2 28 01	12.288
102	1 04 04	10.100	152	2 31 04	12.329
103	1 06 09	10.149	153	2 34 09	12.369
104	1 08 16	10.198	154	2 37 16	12.410
105	1 10 25	10.247	155	2 40 25	12.450
106	1 12 36	10.296	156	2 43 36	12.490
107	1 14 49	10.344	157	2 46 49	12.530
108	1 16 64	10.392	158	2 49 64	12.570
109	1 18 81	10.440	159	2 52 81	12.610
110	1 21 00	10.488	160	2 56 00	12.649
111	1 23 21	10.536	161	2 59 21	12.689
112	1 25 44	10.583	162	2 62 44	12.728
113	1 27 69	10.630	163	2 65 69	12.767
114	1 29 96	10.677	164	2 68 96	12.806
115	1 32 25	10.724	165	2 72 25	12.845
116	1 34 56	10.770	166	2 75 56	12.884
117	1 36 89	10.817	167	2 78 89	12.923
118	1 39 24	10.863	168	2 82 24	12.961
119	1 41 61	10.909	169	2 85 61	13.000
120	1 44 00	10.954	170	2 89 00	13.038
121	1 46 41	11.000	171	2 92 41	13.077
122	1 48 84	11.045	172	2 95 84	13.115
123	1 51 29	11.091	173	2 99 29	13.153
124	1 53 76	11.136	174	3 02 76	13.191
125	1 56 25	11.180	175	3 06 25	13.229
126	1 58 76	11.225	176	3 09 76	13.266
127	1 61 29	11.269	177	3 13 29	13.304
128	1 63 84	11.314	178	3 16 84	13.342
129	1 66 41	11.358	179	3 20 41	13.379
130	1 69 00	11.402	180	3 24 00	13.416
131	1 71 61	11.446	181	3 27 61	13.454
132	1 74 24	11.489	182	3 31 24	13.491
133	1 76 89	11.533	183	3 34 89	13.528
134	1 79 56	11.576	184	3 38 56	13.565
135	1 82 25	11.619	185	3 42 25	13.601
136	1 84 96	11.662	186	3 45 96	13.638
137	1 87 69	11.705	187	3 49 69	13.675
138	1 90 44	11.747	188	3 53 44	13.711
139	1 93 21	11.790	189	3 57 21	13.748
140	1 96 00	11.832	190	3 61 00	13.784
141	1 98 81	11.874	191	3 64 81	13.820
142	2 01 64	11.916	192	3 68 64	13.856
143	2 04 49	11.958	193	3 72 49	13.892
144	2 07 36	12.000	194	3 76 36	13.928
145	2 10 25	12.042	195	3 80 25	13.964
146	2 13 16	12.083	196	3 84 16	14.000
147	2 16 09	12.124	197	3 88 09	14.036
148	2 19 04	12.166	198	3 92 04	14.071
149	2 22 01	12.207	199	3 96 01	14.107
150	2 25 00	12.247	200	4 00 00	14.142

Table of Squares and Square Roots—*Continued*

Number	Square	Square Root	Number	Square	Square Root
201	4 04 01	14.177	251	6 30 01	15.843
202	4 08 04	14.213	252	6 35 04	15.875
203	4 12 09	14.248	253	6 40 09	15.906
204	4 16 16	14.283	254	6 45 16	15.937
205	4 20 25	14.318	255	6 50 25	15.969
206	4 24 36	14.353	256	6 55 36	16.000
207	4 28 49	14.387	257	6 60 49	16.031
208	4 32 64	14.422	258	6 65 64	16.062
209	4 36 81	14.457	259	6 70 81	16.093
210	4 41 00	14.491	260	6 76 00	16.125
211	4 45 21	14.526	261	6 81 21	16.155
212	4 49 44	14.560	262	6 86 44	16.186
213	4 53 69	14.595	263	6 91 69	16.217
214	4 57 96	14.629	264	6 96 96	16.248
215	4 62 25	14.663	265	7 02 25	16.279
216	4 66 56	14.697	266	7 07 56	16.310
217	4 70 89	14.731	267	7 12 89	16.340
218	4 75 24	14.765	268	7 18 24	16.371
219	4 79 61	14.799	269	7 23 61	16.401
220	4 84 00	14.832	270	7 29 00	16.432
221	4 88 41	14.866	271	7 34 41	16.462
222	4 92 84	14.900	272	7 39 84	16.492
223	4 97 29	14.933	273	7 45 29	16.523
224	5 01 76	14.967	274	7 50 76	16.553
225	5 06 25	15.000	275	7 56 25	16.583
226	5 10 76	15.033	276	7 61 76	16.613
227	5 15 29	15.067	277	7 67 29	16.643
228	5 19 84	15.100	278	7 72 84	16.673
229	5 24 41	15.133	279	7 78 41	16.703
230	5 29 00	15.166	280	7 84 00	16.733
231	5 33 61	15.199	281	7 89 61	16.763
232	5 38 24	15.232	282	7 95 24	16.793
233	5 42 89	15.264	283	8 00 89	16.823
234	5 47 56	15.297	284	8 06 56	16.852
235	5 52 25	15.330	285	8 12 25	16.882
236	5 56 96	15.362	286	8 17 96	16.912
237	5 61 69	15.395	287	8 23 69	16.941
238	5 66 44	15.427	288	8 29 44	16.971
239	5 71 21	15.460	289	8 35 21	17.000
240	5 76 00	15.492	290	8 41 00	17.029
241	5 80 81	15.524	291	8 46 81	17.059
242	5 85 64	15.556	292	8 52 64	17.088
243	5 90 49	15.588	293	8 58 49	17.117
244	5 95 36	15.620	294	8 64 36	17.146
245	6 00 25	15.652	295	8 70 25	17.176
246	6 05 16	15.684	296	8 76 16	17.205
247	6 10 09	15.716	297	8 82 09	17.234
248	6 15 04	15.748	298	8 88 04	17.263
249	6 20 01	15.780	299	8 94 01	17.292
250	6 25 00	15.811	300	9 00 00	17.321

TABLE OF SQUARES AND SQUARE ROOTS—*Continued*

Number	Square	Square Root	Number	Square	Square Root
301	9 06 01	17.349	351	12 32 01	18.735
302	9 12 04	17.378	352	12 39 04	18.762
303	9 18 09	17.407	353	12 46 09	18.788
304	9 24 16	17.436	354	12 53 16	18.815
305	9 30 25	17.464	355	12 60 25	18.841
306	9 36 36	17.493	356	12 67 36	18.868
307	9 42 49	17.521	357	12 74 49	18.894
308	9 48 64	17.550	358	12 81 64	18.921
309	9 54 81	17.578	359	12 88 81	18.947
310	9 61 00	17.607	360	12 96 00	18.974
311	9 67 21	17.635	361	13 03 21	19.000
312	9 73 44	17.664	362	13 10 44	19.026
313	9 79 69	17.692	363	13 17 69	19.053
314	9 85 96	17.720	364	13 24 96	19.079
315	9 92 25	17.748	365	13 32 25	19.105
316	9 98 56	17.776	366	13 39 56	19.131
317	10 04 89	17.804	367	13 46 89	19.157
318	10 11 24	17.833	368	13 54 24	19.183
319	10 17 61	17.861	369	13 61 61	19.209
320	10 24 00	17.889	370	13 69 00	19.235
321	10 30 41	17.916	371	13 76 41	19.261
322	10 36 84	17.944	372	13 83 84	19.287
323	10 43 29	17.972	373	13 91 29	19.313
324	10 49 76	18.000	374	13 98 76	19.339
325	10 56 25	18.028	375	14 06 25	19.363
326	10 62 76	18.055	376	14 13 76	19.391
327	10 69 29	18.083	377	14 21 29	19.416
328	10 75 84	18.111	378	14 28 84	19.442
329	10 82 41	18.138	379	14 36 41	19.468
330	10 89 00	18.166	380	14 44 00	19.494
331	10 95 61	18.193	381	14 51 61	19.519
332	11 02 24	18.221	382	14 59 24	19.545
333	11 08 89	18.248	383	14 66 89	19.570
334	11 15 56	18.276	384	14 74 56	19.596
335	11 22 25	18.303	385	14 82 25	19.621
336	11 28 96	18.330	386	14 89 96	19.647
337	11 35 69	18.358	387	14 97 69	19.672
338	11 42 44	18.385	388	15 05 44	19.698
339	11 49 21	18.412	389	15 13 21	19.723
340	11 56 00	18.439	390	15 21 00	19.748
341	11 62 81	18.466	391	15 28 81	19.774
342	11 69 64	18.493	392	15 36 64	19.799
343	11 76 49	18.520	393	15 44 49	19.824
344	11 83 36	18.547	394	15 52 36	19.849
345	11 90 25	18.574	395	15 60 25	19.875
346	11 97 16	18.601	396	15 68 16	19.900
347	12 04 09	18.628	397	15 76 09	19.925
348	12 11 04	18.655	398	15 84 04	19.950
349	12 18 01	18.682	399	15 92 01	19.975
350	12 25 00	18.708	400	16 00 00	20.000

TABLE OF SQUARES AND SQUARE ROOTS—*Continued*

Number	Square	Square Root	Number	Square	Square Root
401	16 08 01	20.025	451	20 34 01	21.237
402	16 16 04	20.050	452	20 43 04	21.260
403	16 24 09	20.075	453	20 52 09	21.284
404	16 32 16	20.100	454	20 61 16	21.307
405	16 40 25	20.125	455	20 70 25	21.331
406	16 48 36	20.149	456	20 79 36	21.354
407	16 56 49	20.174	457	20 88 49	21.378
408	16 64 64	20.199	458	20 97 64	21.401
409	16 72 81	20.224	459	21 06 81	21.424
410	16 81 00	20.248	460	21 16 00	21.448
411	16 89 21	20.273	461	21 25 21	21.471
412	16 97 44	20.298	462	21 34 44	21.494
413	17 05 69	20.322	463	21 43 69	21.517
414	17 13 96	20.347	464	21 52 96	21.541
415	17 22 25	20.372	465	21 62 25	21.564
416	17 30 56	20.396	466	21 71 56	21.587
417	17 38 89	20.421	467	21 80 89	21.610
418	17 47 24	20.445	468	21 90 24	21.633
419	17 55 61	20.469	469	21 99 61	21.656
420	17 64 00	20.494	470	22 09 00	21.679
421	17 72 41	20.518	471	22 18 41	21.703
422	17 80 84	20.543	472	22 27 84	21.726
423	17 89 29	20.567	473	22 37 29	21.749
424	17 97 76	20.591	474	22 46 76	21.772
425	18 06 25	20.616	475	22 56 25	21.794
426	18 14 76	20.640	476	22 65 76	21.817
427	18 23 29	20.664	477	22 75 29	21.840
428	18 31 84	20.688	478	22 84 84	21.863
429	18 40 41	20.712	479	22 94 41	21.886
430	18 49 00	20.736	480	23 04 00	21.909
431	18 57 61	20.761	481	23 13 61	21.932
432	18 66 24	20.785	482	23 23 24	21.954
433	18 74 89	20.809	483	23 32 89	21.977
434	18 83 56	20.833	484	23 42 56	22.000
435	18 92 25	20.857	485	23 52 25	22.023
436	19 00 96	20.881	486	23 61 96	22.045
437	19 09 69	20.905	487	23 71 69	22.068
438	19 18 44	20.928	488	23 81 44	22.091
439	19 27 21	20.952	489	23 91 21	22.113
440	19 36 00	20.976	490	24 01 00	22.136
441	19 44 81	21.000	491	24 10 81	22.159
442	19 53 64	21.024	492	24 20 64	22.181
443	19 62 49	21.048	493	24 30 49	22.204
444	19 71 36	21.071	494	24 40 36	22.226
445	19 80 25	21.095	495	24 50 25	22.249
446	19 89 16	21.119	496	24 60 16	22.271
447	19 98 09	21.142	497	24 70 09	22.293
448	20 07 04	21.166	498	24 80 04	22.316
449	20 16 01	21.190	499	24 90 01	22.338
450	20 25 00	21.213	500	25 00 00	22.361

TABLE OF SQUARES AND SQUARE ROOTS—*Continued*

Number	Square	Square Root	Number	Square	Square Root
501	25 10 01	22.383	551	30 36 01	23.473
502	25 20 04	22.405	552	30 47 04	23.495
503	25 30 09	22.428	553	30 58 09	23.516
504	25 40 16	22.450	554	30 69 16	23.537
505	25 50 25	22.472	555	30 80 25	23.558
506	25 60 36	22.494	556	30 91 36	23.580
507	25 70 49	22.517	557	31 02 49	23.601
508	25 80 64	22.539	558	31 13 64	23.622
509	25 90 81	22.561	559	31 24 81	23.643
510	26 01 00	22.583	560	31 36 00	23.664
511	26 11 21	22.605	561	31 47 21	23.685
512	26 21 44	22.627	562	31 58 44	23.707
513	26 31 69	22.650	563	31 69 69	23.728
514	26 41 96	22.672	564	31 80 96	23.749
515	26 52 25	22.694	565	31 92 25	23.770
516	26 62 56	22.716	566	32 03 56	23.791
517	26 72 89	22.738	567	32 14 89	23.812
518	26 83 24	22.760	568	32 26 24	23.833
519	26 93 61	22.782	569	32 37 61	23.854
520	27 04 00	22.804	570	32 49 00	23.875
521	27 14 41	22.825	571	32 60 41	23.896
522	27 24 84	22.847	572	32 71 84	23.917
523	27 35 29	22.869	573	32 83 29	23.937
524	27 45 76	22.891	574	32 94 76	23.958
525	27 56 25	22.913	575	33 06 25	23.979
526	27 66 76	22.935	576	33 17 76	24.000
527	27 77 29	22.956	577	33 29 29	24.021
528	27 87 84	22.978	578	33 40 84	24.042
529	27 98 41	23.000	579	33 52 41	24.062
530	28 09 00	23.022	580	33 64 00	24.083
531	28 19 61	23.043	581	33 75 61	24.104
532	28 30 24	23.065	582	33 87 24	24.125
533	28 40 89	23.087	583	33 98 89	24.145
534	28 51 56	23.108	584	34 10 56	24.166
535	28 62 25	23.130	585	34 22 25	24.187
536	28 72 96	23.152	586	34 33 96	24.207
537	28 83 69	23.173	587	34 45 69	24.228
538	28 94 44	23.195	528	34 57 44	24.249
539	29 05 21	23.216	589	34 69 21	24.269
540	29 16 00	23.238	590	34 81 00	24.290
541	29 26 81	23.259	591	34 92 81	24.310
542	29 37 64	23.281	592	35 04 64	24.331
543	29 48 49	23.302	593	35 16 49	24.352
544	29 59 36	23.324	594	35 28 36	24.372
545	29 70 25	23.345	595	35 40 25	24.393
546	29 81 16	23.367	596	35 52 16	24.413
547	29 92 09	23.388	597	35 64 09	24.434
548	30 03 04	23.409	598	35 76 04	24.454
549	30 14 01	23.431	599	35 88 01	24.474
550	30 25 00	23.452	600	36 00 00	24.495

TABLE OF SQUARES AND SQUARE ROOTS—*Continued*

Number	Square	Square Root	Number	Square	Square Root
601	36 12 01	24.515	651	42 38 01	25.515
602	36 24 04	24.536	652	42 51 04	25.534
603	36 36 09	24.556	653	42 64 09	25.554
604	36 48 16	24.576	654	42 77 16	25.573
605	36 60 25	24.597	655	42 90 25	25.593
606	36 72 36	24.617	656	43 03 36	25.612
607	36 84 49	24.637	657	43 16 49	25.632
608	36 96 64	24.658	658	43 29 64	25.652
609	37 08 81	24.678	659	43 42 81	25.671
610	37 21 00	24.698	660	43 56 00	25.690
611	37 33 21	24.718	661	43 69 21	25.710
612	37 45 44	24.739	662	43 82 44	25.729
613	37 57 69	24.759	663	43 95 69	25.749
614	37 69 96	24.779	664	44 08 96	25.768
615	37 82 25	24.799	665	44 22 25	25.788
616	37 94 56	24.819	666	44 35 56	25.807
617	38 06 89	24.839	667	44 48 89	25.826
618	38 19 24	24.860	668	44 62 24	25.846
619	38 31 61	24.880	669	44 75 61	25.865
620	38 44 00	24.900	670	44 89 00	25.884
621	38 56 41	24.920	671	45 02 41	25.904
622	38 68 84	24.940	672	45 15 84	25.923
623	38 81 29	24.960	673	45 29 29	25.942
624	38 93 76	24.980	674	45 42 76	25.962
625	39 06 25	25.000	675	45 56 25	25.981
626	39 18 76	25.020	676	45 69 76	26.000
627	39 31 29	25.040	677	45 83 29	26.019
628	39 43 84	25.060	678	45 96 84	26.038
629	39 56 41	25.080	679	46 10 41	26.058
630	39 69 00	25.100	680	46 24 00	26.077
631	39 81 61	25.120	681	46 37 61	26.096
632	39 94 24	25.140	682	46 51 24	26.115
633	40 06 89	25.159	683	46 64 89	26.134
634	40 19 56	25.179	684	46 78 56	26.153
635	40 32 25	25.199	685	46 92 25	26.173
636	40 44 96	25.219	686	47 05 96	26.192
637	40 57 69	25.239	687	47 19 69	26.211
638	40 70 44	25.259	688	47 33 44	26.230
639	40 83 21	25.278	689	47 47 21	26.249
640	40 96 00	25.298	690	47 61 00	26.268
641	41 08 81	25.318	691	47 74 81	26.287
642	41 21 64	25.338	692	47 88 64	26.306
643	41 34 49	25.357	693	48 02 49	26.325
644	41 47 36	25.377	694	48 16 36	26.344
645	41 60 25	25.397	695	48 30 25	26.363
646	41 73 16	25.417	696	48 44 16	26.382
647	41 86 09	25.436	697	48 58 09	26.401
648	41 99 04	25.456	698	48 72 04	26.420
649	42 12 01	25.475	699	48 86 01	26.439
650	42 25 00	25.495	700	49 00 00	26.458

Table of Squares and Square Roots—*Continued*

Number	Square	Square Root	Number	Square	Square Root
701	49 14 01	26.476	751	56 40 01	27.404
702	49 28 04	26.495	752	56 55 04	27.423
703	49 42 09	26.514	753	56 70 09	27.441
704	49 56 16	26.533	754	56 85 16	27.459
705	49 70 25	26.552	755	57 00 25	27.477
706	49 84 36	26.571	756	57 15 36	27.495
707	49 98 49	26.589	757	57 30 49	27.514
708	50 12 64	26.608	758	57 45 64	27.532
709	50 26 81	26.627	759	57 60 81	27.550
710	50 41 00	26.646	760	57 76 00	27.568
711	50 55 21	26.665	761	57 91 21	27.586
712	50 69 44	26.683	762	58 06 44	27.604
713	50 83 69	26.702	763	58 21 69	27.622
714	50 97 96	26.721	764	58 36 96	27.641
715	51 12 25	26.739	765	58 52 25	27.659
716	51 26 56	26.758	766	58 67 56	27.677
717	51 40 89	26.777	767	58 82 89	27.695
718	51 55 24	26.796	768	58 98 24	27.713
719	51 69 61	26.814	769	59 13 61	27.731
720	51 84 00	26.833	770	59 29 00	27.749
721	51 98 41	26.851	771	59 44 41	27.767
722	52 12 84	26.870	772	59 59 84	27.785
723	52 27 29	26.889	773	59 75 29	27.803
724	52 41 76	26.907	774	59 90 76	27.821
725	52 56 25	26.926	775	60 06 25	27.839
726	52 70 76	26.944	776	60 21 76	27.857
727	52 85 29	26.963	777	60 37 29	27.875
728	52 99 84	26.981	778	60 52 84	27.893
729	53 14 41	27.000	779	60 68 41	27.911
730	53 29 00	27.019	780	60 84 00	27.928
731	53 43 61	27.037	781	60 99 61	27.946
732	53 58 24	27.055	782	61 15 24	27.964
733	53 72 89	27.074	783	61 30 89	27.982
734	53 87 56	27.092	784	61 46 56	28.000
735	54 02 25	27.111	785	61 62 25	28.018
736	54 16 96	27.129	786	61 77 96	28.036
737	54 31 69	27.148	787	61 93 69	28.054
738	54 46 44	27.166	788	62 09 44	28.071
739	54 61 21	27.185	789	62 25 21	28.089
740	54 76 00	27.203	790	62 41 00	28.107
741	54 90 81	27.221	791	62 56 81	28.125
742	55 05 64	27.240	792	62 72 64	28.142
743	55 20 49	27.258	793	62 88 49	28.160
744	55 35 36	27.276	794	63 04 36	28.178
745	55 50 25	27.295	795	63 20 25	28.196
746	55 65 16	27.313	796	63 36 16	28.213
747	55 80 09	27.331	797	63 52 09	28.231
748	55 95 04	27.350	798	63 68 04	28.249
749	56 10 01	27.368	799	63 84 01	28.267
750	56 25 00	27.386	800	64 00 00	28.284

Table of Squares and Square Roots—*Continued*

Number	Square	Square Root	Number	Square	Square Root
801	64 16 01	28.302	851	72 42 01	29.172
802	64 32 04	28.320	852	72 59 04	29.189
803	64 48 09	28.337	853	72 76 09	29.206
804	64 64 16	28.355	854	72 93 16	29.223
805	64 80 25	28.373	855	73 10 25	29.240
806	64 96 36	28.390	856	73 27 36	29.257
807	65 12 49	28.408	857	73 44 49	29.275
808	65 28 64	28.425	858	73 61 64	29.292
809	65 44 81	28.443	859	73 78 81	29.309
810	65 61 00	28.460	860	73 96 00	29.326
811	65 77 21	28.478	861	74 13 21	29.343
812	65 93 44	28.496	862	74 30 44	29.360
813	66 09 69	28.513	863	74 47 69	29.377
814	66 25 96	28.531	864	74 64 96	29.394
815	66 42 25	28.548	865	74 82 25	29.411
816	66 58 56	28.566	866	74 99 56	29.428
817	66 74 89	28.583	867	75 16 89	29.445
818	66 91 24	28.601	868	75 34 24	29.462
819	67 07 61	28.618	869	75 51 61	29.479
820	67 24 00	28.636	870	75 69 00	29.496
821	67 40 41	28.653	871	75 86 41	29.513
822	67 56 84	28.671	872	76 03 84	29.530
823	67 73 29	28.688	873	76 21 29	29.547
824	67 89 76	28.705	874	76 38 76	29.563
825	68 06 25	28.723	875	76 56 25	29.580
826	68 22 76	28.740	876	76 73 76	29.597
827	68 39 29	28.758	877	76 91 29	29.614
828	68 55 84	28.775	878	77 08 84	29.631
829	68 72 41	28.792	879	77 26 41	29.648
830	68 89 00	28.810	880	77 44 00	29.665
831	69 05 61	28.827	881	77 61 61	29.682
832	69 22 24	28.844	882	77 79 24	29.698
833	69 38 89	28.862	883	77 96 89	29.715
834	69 55 56	28.879	884	78 14 56	29.732
835	69 72 25	28.896	885	78 32 25	29.749
836	69 88 96	28.914	886	78 49 96	29.766
837	70 05 69	28.931	887	78 67 69	29.783
838	70 22 44	28.948	888	78 85 44	29.799
839	70 39 21	28.965	889	79 03 21	29.816
840	70 56 00	28.983	890	79 21 00	29.833
841	70 72 81	29.000	891	79 38 81	29.850
842	70 89 64	29.017	892	79 56 64	29.866
843	71 06 49	29.034	893	79 74 49	29.883
844	71 23 36	29.052	894	79 92 36	29.900
845	71 40 25	29.069	895	80 10 25	29.916
846	71 57 16	29.086	896	80 28 16	29.933
847	71 74 09	29.103	897	80 46 09	29.950
848	71 91 04	29.120	898	80 64 04	29.967
849	72 08 01	29.138	899	80 82 01	29.983
850	72 25 00	29.155	900	81 00 00	30.000

Table of Squares and Square Roots—*Continued*

Number	Square	Square Root	Number	Square	Square Root
901	81 18 01	30.017	951	90 44 01	30.838
902	81 36 04	30.033	952	90 63 04	30.854
903	81 54 09	30.050	953	90 82 09	30.871
904	81 72 16	30.067	954	91 01 16	30.887
905	81 90 25	30.083	955	91 20 25	30.903
906	82 08 36	30.100	956	91 39 36	30.919
907	82 26 49	30.116	957	91 58 49	30.935
908	82 44 64	30.133	958	91 77 64	30.952
909	82 62 81	30.150	959	91 96 81	30.968
910	82 81 00	30.166	960	92 16 00	30.984
911	82 99 21	30.183	961	92 35 21	31.000
912	83 17 44	30.199	962	92 54 44	31.016
913	83 35 69	30.216	963	92 73 69	31.032
914	83 53 96	30.232	964	92 92 96	31.048
915	83 72 25	30.249	965	93 12 25	31.064
916	83 90 56	30.265	966	93 31 56	31.081
917	84 08 89	30.282	967	93 50 89	31.097
918	84 27 24	30.299	968	93 70 24	31.113
919	84 45 61	30.315	969	93 89 61	31.129
920	84 64 00	30.332	970	94 09 00	31.145
921	84 82 41	30.348	971	94 28 41	31.161
922	85 00 84	30.364	972	94 47 84	31.177
923	85 19 29	30.381	973	94 67 29	31.193
924	85 37 76	30.397	974	94 86 76	31.209
925	85 56 25	30.414	975	95 06 25	31.225
926	85 74 76	30.430	976	95 25 76	31.241
927	85 93 29	30.447	977	95 45 29	31.257
928	86 11 84	30.463	978	95 64 84	31.273
929	86 30 41	30.480	979	95 84 41	31.289
930	86 49 00	30.496	980	96 04 00	31.305
931	86 67 61	30.512	981	96 23 61	31.321
932	86 86 24	30.529	982	96 43 24	31.337
933	87 04 89	30.545	983	96 62 89	31.353
934	87 23 56	30.561	984	96 82 56	31.369
935	87 42 25	30.578	985	97 02 25	31.385
936	87 60 96	30.594	986	97 21 96	31.401
937	87 79 69	30.610	987	97 41 69	31.417
938	87 98 44	30.627	988	97 61 44	31.432
939	88 17 21	30.643	989	97 81 21	31.448
940	88 36 00	30.659	990	98 01 00	31.464
941	88 54 81	30.676	991	98 20 81	31.480
942	88 73 64	30.692	992	98 40 64	31.496
943	88 92 49	30.708	993	98 60 49	31.512
944	89 11 36	30.725	994	98 80 36	31.528
945	89 30 25	30.741	995	99 00 25	31.544
946	89 49 16	30.757	996	99 20 16	31.559
947	89 68 09	30.773	997	99 40 09	31.575
948	89 87 04	30.790	998	99 60 04	31.591
949	90 06 01	30.806	999	99 80 01	31.607
950	90 25 00	30.822	1000	100 00 00	31.623

INDEX

Italics are used for Reference to Definitions.

HA33 c. 1
G192 Garrett, Henry Edward.
Statistics in Psychology and
education.

Date Due			
~~7-30-71~~			

HA33 c. 1
G192 Garrett, Henry Edward.

AUTHOR Statistics in psychology
and Education.

TITLE

DATE DUE	BORROWER'S NAME
	Reference Copy
~~7-30-71~~	~~L Barron~~